Commentary

MAN AND HIS ENVIRONMENT

THE RECENT presidential address delivered by Dr L. Harrison Matthews, F.R.S., to the Zoology Section of the British Association emphasized again some urgently needed developments in biological science, and in the use to be made of the results of biological research. When the human population of the world was relatively small, man was probably in a stable relationship with his environment; he preyed upon the fauna, but did not in the long run deplete it. The steady increase of numbers, and the development of human technical skills have, however, upset this stable relationship, and resulted in considerable destruction of the environment. Man is entirely dependent for food, and for many other materials, upon living things, and continued destruction of these and of their habitats can only have the most disastrous consequences. A reappraisal of man's approach to his livelihood is long overdue.

Fundamentally the problem starts with the vast increase of human numbers, particularly in the less well developed countries. During 1958, according to the United States Population Reference Bureau, the total population of the world increased by 47 millions. At this rate the world population would be doubled in about 40 years, and in just over 1000 years there would be one person for every square yard of the earth's surface. This is, of course, unthinkable. Something will act to limit the population—but what? War? Nuclear weapons have certainly conferred upon man the power to render his species very sparse indeed. We can only hope he will spare himself this ultimate insanity. Pestilence? Decimation of the human race by disease is not so unlikely as it sounds, particularly since stress disease appears to be increasing in America and Europe. This syndrome is similar to the one which causes dramatic reductions in populations of animals, such as voles, lemmings and jack-rabbits, which indulge in uncontrolled population increases.

A third possibility exists, however. Limitation of reproduction would relieve the awful pressure, and enable the human race to stabilize its numbers at a reasonable level. There seems little hope that the human species will take this sane and sensible step in the near future—religious and political pressure remains too strong. It is sometimes argued that, since increased standards of living result in reduced rates of population growth, there is no need to take special steps to limit populations. Unfortunately, it seems unlikely that the increasing world population will stabilize soon enough, or at a low enough figure (bearing in mind that the stress syndrome is a direct result of overcrowding) to solve the problem.

While we wait for humanity to mature sufficiently to accept responsibility for its ultimate fate, we must ensure that there is enough food for the newborn millions. This means a sharp revision of policy in the exploitation of the resources of both land and sea, because between 1939 and 1958, according to the United States Department of Agriculture, food production in the 'free world' increased by only one or two per cent per head, while in Communist countries it *declined* by nearly seven per cent per head. To remedy this disastrous situation, two major steps are needed—a vast increase of biological research, and a much more determined use by governments of the results of this work.

Biology has much to offer. Under careful management based on adequate knowledge the elephant seal and the fur seal yield large crops yet continue to increase. The biological information exists for similar profitable management of the whale fishery, but political disagreement prevents its being put properly into effect, so that the whaling industry seems well on the road to destroying itself, and many species of whales in the process. Legislation has so far failed to prevent overfishing of the North Sea, although marine biologists have made clear many of the reasons for the continued low productivity of this stretch of water.

On land, there must be an end to the limited thinking which can see only traditional pastoralism and agriculture when faced with the job of developing a new territory. In many cases thorough ecological investigation might reveal some way of exploiting the truly natural resources of a region, where clearing and planting would shortly add another desolate dust bowl to the dismal score of man's mistakes. For example, Darling has pointed out that the indigenous African ungulates, living in delicate adjustment with their poor and tender habitat, can convert the vegetation into valuable protein far more efficiently than any imported cattle, sheep or goats, so that replacement of the former by the latter is sheer waste. Even in lands where agriculture is well established, this new concept of conservation and land management has a valuable contribution to make. If governments are to do their proper duty to their peoples, they must invest greatly increased sums of money in fundamental ecological investigation, and in the development of methods of conserving and exploiting, rather than destroying, the world's natural living resources. I.J.L.

RADIOISOTOPES IN THE DATING OF GEOLOGICAL AND ARCHAEOLOGICAL EVENTS

B. J. GILETTI AND R. St J. LAMBERT

Department of Geology and Mineralogy, University Museum, Oxford

Methods for determining the age of geological and archaeological events by the use of radio isotopes are reviewed. The C^{14} method is described together with its application to events up to 70,000 years old. The U–Th–Pb, K^{40}–A^{40}, and Rb^{87}–Sr^{87} decay schemes are discussed in terms of their application to dating geological events which may be millions or even thousands of millions of years old.

In such studies as archaeology and geology, the relative order and rates at which events take place are extremely important. It is often impossible, however, to establish the contemporaneity or relative ages of physically separated events without accurate absolute dates.

Until recently, the most reliable dating methods employed stratigraphical principles. The establishment of a sequence of fossils or artifacts based on the superposition of younger strata on older yielded a satisfactory relative time scale. Often, however, guide fossils or artifacts are lacking or limited in geographical extent, and chronologies cannot be established. Some absolute methods are therefore required which can date either geologically ancient or the more recent archaeological–geological events. Even though the principle of dating rocks by analysis for a radioactive element and its decay product has been known almost as long as radioactivity itself, no reliable methods involving specific isotopes were known until 1938, when Nier observed variations in the isotopic composition of common leads. He correctly inferred that these were caused by the addition of lead, from the decay of uranium and thorium, to the rocks from which lead ores are generated. From this beginning, methods have been developed whereby minerals that are rich in a radioactive nuclide can be dated. *Table 1* lists the more commonly used nuclides. In each case the parent: daughter ratio is obtained and the time since the crystal formed is calculated using the radioactive decay law, assuming that the mineral was a closed system during the whole of its existence. These methods date the very ancient events, while the carbon 14 method covers the more recent ages.

Dating the Recent Past

It is only about ten years since W. F. Libby and his co-workers established the C^{14} method for dating materials which are up to several tens of thousands of years old[1]. Carbon 14 is beta radioactive with a half life of 5568 years. It is produced by cosmic ray reactions in the upper atmosphere, oxidized, and rapidly distributed throughout the atmosphere as carbon dioxide, whence it is incorporated into plant and animal tissue. The rate of C^{14} production, and also the C^{14} concentration in living organisms, are assumed to have been constant during the last 70 thousand years.

Once the organism dies, however, it becomes a closed system no longer in equilibrium with the C^{14} of the atmosphere. The present radioactivity of a dead organism is that fraction of the activity at time of death determined by the length of time since it died. Thus, cloth made from cotton grown 5568 years ago would have one half the activity of cloth made from this year's crop. In principle, then, it is only necessary to determine the present day equilibrium C^{14} concentration in living organisms and to compare this with that in the sample to be dated. No daughter product is involved.

Materials which can be dated include wood, charcoal, leaves, flesh, hair, horn, bone, hide, parchment and shell. Laboratories vary in requirement, but usually need two to ten grams of carbon. Owing to its minute concentration (10^{-12} g of carbon 14 per gram of carbon) C^{14} is best determined by radioactive counting. Among the methods currently in use are: solid source (carbon black) counted inside a geiger counter; internal gas proportional counting (using carbon dioxide,

acetylene or methane); liquid scintillation counting (using toluene or liquid carbon dioxide). The most widely used technique employs internal gas proportional counting.

The preparation begins with the conversion of the sample to carbon dioxide by either oxidation or reaction with acid. This forms the starting point either for purification and direct counting or for chemical conversion to carbon, acetylene, methane, or toluene, depending on the laboratory.

Because of the low energy of the C^{14} beta particle, and the low specific activity of even modern samples (about thirteen disintegrations per gram carbon per minute), highly sensitive counters with low, stable backgrounds are required. Large counters (two to seven litres in volume) which can be operated at two, three, or even five atmospheres of sample gas pressure are constructed to increase sample size. Larger samples result in increased sensitivity and each factor of two gained in sensitivity extends the dateable age back one half life (5568 years).

Another means of increasing sensitivity is by the reduction and stabilization of the counter background. These large counters normally have backgrounds of several thousand counts per minute. Several inches of lead and/or steel shielding reduce this rate to several hundred counts per minute. A sample which is 30,000 years old will contribute only about one count per minute, using an average counter, and the background would obscure this.

To reduce the counter background further, a technique called anti-coincidence counting is employed. Most of the residual background is from high energy mesons which cannot be stopped even by elaborate shielding. A ring of Geiger counters is constructed to surround the sample counter closely. In order for a meson to reach the sample counter it must also pass through this ring of geiger tubes, called an anti-coincidence or A-C ring. The electronics are then so arranged that whenever a signal is received simultaneously (in coincidence) from sample counter and A-C ring, no count is recorded for the sample counter. In this way several hundred meson counts per minute are subtracted from the background, and the net background will then be only ten to twenty counts per minute. The C^{14} betas are too weak to reach the A-C counters and so are not detected by the A-C ring at all.

With these precautions, it is possible to date objects as old as 40,000 years. Recently, still more elaborate shielding, such as borax–paraffin wax mixtures for neutron shielding, mercury for shielding off the gamma and x-rays from the steel or lead and A-C rings incorporated into the sample counter itself, has been devised so that it is possible to date 60,000 year old objects.

The reliability of the C^{14} method has been tested by dating materials whose age is known by independent means. *Figure 1* shows that this comparison renders the C^{14} technique valid.

Age Determinations using Isotopes of Long Half Life

It has been found that the age of the earth's crust is such that isotopes of half lives in the range 10^8 to 10^{11} years may be used to date geological events. Suitable isotopes are U^{238}, U^{235}, Th^{232}, Rb^{87} and K^{40} (see *Table 1*). Others such as Re^{187} are too rare to be commonly applicable.

The Uranium–Thorium–Lead Method—This is the method of greatest antiquity but not of the greatest promise. A mineral containing appreciable uranium or thorium is analysed for total uranium, thorium and lead, and the lead is analysed isotopically. The age is then found from the following equation:

$$\frac{\text{daughter isotope}}{\text{parent isotope}} = e^{\lambda t} - 1$$

where λ is the decay constant and t is the age.

Table 1. Naturally occurring radionuclides

Nuclide	*Half life* (years)	*Abundance* (% of element)	*Useful age range* (million years)	*Occurrence*
K^{40}	β–decay $1{\cdot}47\times10^{9}$ κ–capture $1{\cdot}19\times10^{10}$	0·0119	>0·1	Widespread in all rock types: biotite and muscovite micas are most useful.
Rb^{87}	$5{\cdot}0\times10^{10}$	27·85	>100	Rb a relatively rare element: micas most useful.
Th^{232}	$1{\cdot}39\times10^{10}$	100	>10	U and Th occur in local concentrations only. Most useful minerals: uraninite, pitchblende, zircon.
U^{235}	$7{\cdot}13\times10^{8}$	0·710		
U^{238}	$4{\cdot}51\times10^{9}$	99·285		

In principle the method is excellent as the decay constants are well known, internal checks are possible, and the concentrations of parent and daughter can be measured with considerable accuracy. Often, a mineral containing uranium contains thorium also and vice versa. Consequently three independent ages for a particular mineral can be found, *viz.* from the ratios Pb^{206}:U^{238}, Pb^{207}:U^{235}, and Pb^{208}:Th^{232}. The ratio Pb^{207}:Pb^{206} also gives an age, but it is not independent of the others. These ages are 'concordant' if agreeing or 'discordant' otherwise, for a given specimen.

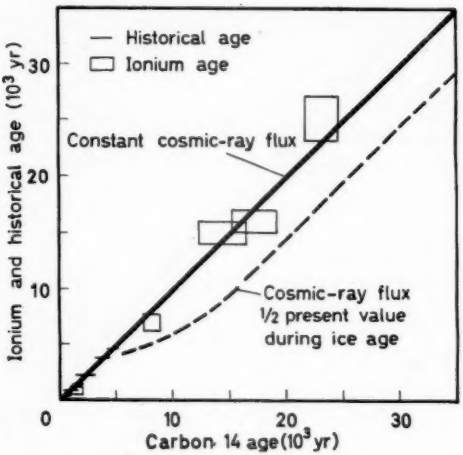


Figure 1. Comparison of dates found using C^{14} technique with known ages (Reproduced by permission of the editor of Nucleonics, (1954) 12, 19)

Uranium may be determined gravimetrically, by polarography, by colorimetry using dibenzoylmethane[2], or by isotope dilution using U^{235} as tracer[3] if present at low concentrations. Total lead may be determined gravimetrically, polarographically, colorimetrically, or by isotope dilution using enriched Pb^{208}. The isotopic analysis of lead may be made using the volatile lead tetramethyl or lead iodide in a gas source mass spectrometer, or lead sulphide or sulphate in a solid source mass spectrometer. The latter method is preferable in the case of small samples.

The presence of lead ('common lead') incorporated in the mineral at the time of crystallization will give too high an age value. The abundance of Pb^{204} can be used to correct for common lead since this isotope is not produced by any known decay scheme. The isotopic constitution of lead varies with time, so lead from a uranium-free mineral associated with the uranium is analysed to supply the correction. If this correction is large, the uncertainty in the ages may be appreciable.

Suitable minerals are, unfortunately, rare. The best, uraninite, is confined to pegmatites and mineralized zones of a local nature, and is rare even in these. Other primary minerals which have been used are thorite, pitchblende, samarskite, zircon, monazite, brannerite, and columbite. Zircon and monazite are widely distributed as minor constituents of igneous rocks, but are difficult to separate from their matrix and may give discordant ages[4]. Uranium is slightly concentrated in the sediment, black shale, but in only one case, the Swedish Kolm of Upper Cambrian age, has anything like a reasonable age been obtained[5]. The only mineral which frequently gives concordant ages is uraninite and *Table 2* gives a list of such ages. It is generally agreed that concordant ages reflect the age of the crystallization of the mineral (or its last recrystallization if metamorphosed).

The Lead Isotope Method—It has been found that the isotopic composition of lead in lead ores (chiefly galena, PbS) varies in many cases with the age of the ore. If it is assumed that the lead was thoroughly mixed with uranium and thorium until the time of crystallization and was then separated from them, it is possible to calculate an age for the ore deposition. It is sometimes possible to discriminate even between young ores; in the British Isles, Hercynian and Caledonian ores—galena ages of 280 and 440 million years respectively—may be easily distinguished[6].

For purposes of absolute age determination the accuracy of this method is not as great as that of the other methods discussed, so that its chief uses are in the assessment of mineral deposits, and in general geochemical problems such as the age and evolution of the earth's crust.

The Potassium–Argon Method—The development of this method has been entirely achieved in the last decade, yet it is probably the most promising of all. The major advantages are that potassium is an abundant and widely distributed element and that argon, though a gas, is retained in micas to a high degree, greater than 90 per cent in most cases. K^{40} decays both by electron capture to A^{40} and by beta decay to Ca^{40}. The ratio λ_e/λ_β is 0·124 (from the constants given in *Table 1*); hence more Ca^{40} is produced than A^{40}. Unfortunately Ca^{40} is the commonest isotope of calcium and there is generally enough primary calcium present in the mineral to

obscure the radiogenic contribution. The principal exception is sylvine (KCl) a rare mineral found in salt deposits. Although the decay scheme of K^{40} is well known, the decay constants have not yet been measured as accurately as would be desirable.

Potassium minerals are found in most types of igneous and metamorphic rock and in some sediments. The sheet silicates biotite, muscovite, and lepidolite (all micas) occur in rocks of high temperature origin and yield ages generally believed to be within 90 per cent of the true age of crystallization. The potassic sheet silicate glauconite, which crystallizes in some sediments at or near the time of their formation, gives ages which appear to approach the probable age of the sediment as derived from other methods. The framework silicates of the alkali feldspar group (alkali aluminium silicates) yield ages which range between 50 and 100 per cent of the mica age for the same rock. The cause of this discrepancy is generally believed to be the gradual loss of argon from the feldspars, which may be related to their complex order–disorder inversions.

The argon is extracted by heating the mineral *in vacuo* by radiofrequency induction, usually at about 1400°C, which causes fusion. A graphite or molybdenum cup containing the sample acts as the secondary of the generator. Other gases present are extracted by use of cold traps cooled by liquid air, a copper–copper oxide furnace, and a barium, titanium or calcium furnace, and the volume of pure argon finally measured in a McLeod gauge. Alternatively, an exactly known quantity of A^{38} may be added during fusion and the resulting mixture analysed on a mass spectrometer, giving the radiogenic argon. This isotope dilution technique has the advantage that recovery of the argon mixture does not have to be quantitative.

Table 2. Concordant uraninite ages compiled from numerous sources

Locality	*Age in millions of years*			
	$\frac{206}{238}$	$\frac{207}{235}$	$\frac{207}{206}$	$\frac{208}{232}$
Cooke City, Montana	2580	2640	2720	—
Witwatersrand	2220	2160	2130	—
Loolekop, E. Transvaal	2000	2020	2060	1820
Viking Lake, Saskatchewan	1850	1880	1910	1670
Ingersoll Mine, S.D.	1580	1600	1630	1440
Wilberforce, Ont.	1060	1055	1015	1000
Cardiff Township, Ont.	1000	1020	1085	870
Gordonia, S. Africa	1010	1015	1037	1215
Parry Sound, Ont.	994	993	1000	897
Romteland, Norway	890	892	920	900
Biundni, India	735	735	740	935
Morogoro, Tanganyika	610	609	595	920
Shinkolobwe, Belg. Congo	610	610	625	—
Palette Lode, N. Terr., Aust.	500	507	533	—
Bemasoandro, Madagascar	460	450	480	465
Chestnut Flat, N. Carolina	375	380	420	—
Portland, Conn.	268	266	250	239

The Rubidium–Strontium Method—Although it has been known since 1937 that Rb^{87} is radioactive with a long half life, it was not until 1953 that reliable measurements were made of the age of minerals using the Rb–Sr method[7]. Determinations of the half life have given widely varying results: $4{\cdot}3\times10^{10}$ years to $6{\cdot}2\times10^{10}$ years. At the present time, $5{\cdot}0\times10^{10}$ years is generally used[8]. A second difficulty is that rubidium is not an abundant element, and moreover, primary strontium is often present in rubidium-bearing minerals to such an extent that it obscures the radiogenic Sr^{87}.

Rubidium does not form any minerals of its own, but shows a close geochemical coherence with potassium, most rocks having a K/Rb ratio[9] of about 240. Thus micas and alkali feldspars may be used for Rb–Sr determinations; muscovite and lepidolite being particularly valuable. Both Rb^{87} and Sr^{87} are measured using isotope dilution techniques, Rb^{87} and Sr^{86} or Sr^{84} being used as tracers. A solid source mass spectrometric analysis is made after the Rb and Sr have been separated on an ion exchange column. The analytical uncertainty is usually about two to three per cent unless there is a large amount of normal strontium present to obscure the radiogenic Sr^{87}. The method is regarded as being of great promise.

Results and Discussion

Carbon 14—In geologically recent times, world climate has been affected by glacial epochs. C^{14} dating has now established that the last widespread glaciation, which covered large areas of Asia, Europe, and North America, receded about eleven thousand years ago. There is increasing evidence that this latest ice age began about 70 thousand years ago—minor extrapolation from C^{14} dates is necessary. *Table 3* lists some of the critical dates used in establishing these results, showing the widely differing sources from which the data converge on the eleven thousand year date.

Carbon 14 dating also lends itself well to the dating of archaeological remains, especially such relatively common objects as charcoal from hearths. In such ways it has been established with C^{14} that people were already widespread in North America at least ten thousand years ago, but no evidence has yet been found for the existence of very early man there.

The oldest evidence for village life comes from Jarmo, Iraq, where a 6707±320 year date (University of Chicago, No. 113) was obtained. The authenticity of the Dead Sea Scrolls is supported by a date of 1917±200 years for the linen wrapping from the Book of Isaiah (University of Chicago, No. 576).

Table 3. Carbon 14 dates

Sample description	*Age*	*Laboratory*
GEOLOGICAL		
Boundary between cold and warm water sediment types in deep ocean bottom	11,000*	L.G.O.†
Climate change from wet to dry in western United States	11,000*	L.G.O.
Forest buried by last minor ice sheet advance	11,000*	U. Chic.‡
Sea level change seen in Mississippi River delta sediments; melting ice raised to sea level	11,000*	L.G.O.
ARCHAEOLOGICAL		
Jarmo, Iraq	6707±320	U. Chic.
Book of Isaiah	1917±200	U. Chic.
Oldest established human sites, North America	11,000*	Various

* Average of numerous dates
† Lamont Geological Observatory, Columbia University
‡ University of Chicago

Recently, two important changes have occurred which affect the basic assumptions made in the C^{14} method. Since the late nineteenth century increasing amounts of coal and petroleum have been burned with a consequent release of this carbon as carbon dioxide. This 'fossil' carbon is too old to contain C^{14} and, therefore, has the effect of diluting the C^{14} of the atmosphere. A comparison of wood from 1890 with that of today shows that this dilution is now about one per cent; it is called the Suess effect, after its discoverer. In the last five years a second factor has entered the picture, namely the production of C^{14} by nuclear explosions. The C^{14} concentration of the atmosphere has increased by several per cent in this time and incidentally cancels the fossil carbon dilution.

The effect of these two man-made factors on dating old materials, however, is nil. The old samples have presumably been closed systems during the last seventy years. Nevertheless the effect on the contemporary atmosphere, and thus on contemporary standards for comparison with the old samples, requires that the standards be chosen with care.

Long Term Dating Methods—Each of the common long term dating methods has some particular advantages over the others. Uranium and thorium have the best known half lives and can give internal consistency checks. The potassium–argon method employs a very common element, potassium, and can be used both for very old and very young samples, while the rubidium–strontium system is the least subject to subsequent external effects during the mineral's history. The best age determinations are those made by a variety of methods on a variety of minerals all from the same rock: if all agree, there is little doubt that the age represents the formation of this rock. This point and the causes and meaning of discordance have been discussed recently by L. T. ALDRICH and G. W. WETHERILL[10].

Much recent age research has been aimed at examining the formation of mountain chains, which are usually the culmination of a long series of processes producing both an elevated terrain and recrystallization of the rocks at depth. The former high ground is subsequently eroded so that the roots of the mountains are found exposed. These contain the rocks which crystallized during the mountain-building and so can be used to date those events. The development of mountain ranges is a major problem whose cause is often to be sought in events taking place deep within the earth. The earth's crust then responds to these deep-seated stresses by fracturing and wrinkling. Examination of the data so far available shows that there may be a widespread correlation of such events. The map of the world, *Figure 2*, shows the present state of knowledge of the history of the crust, omitting dubious and controversial dates and also all 'ages' obtained by the lead-isotope method. The dates given are those indicating metamorphism, igneous intrusion or the formation of major ore bodies. Prominent are dates grouped around 2600 to 2700, 1800, 1600, 1400, 1000, 600, 480, and 300 to 360 million years. It is too early yet, owing to lack of good data, to ascertain how real is this division into distinct episodes of earth movement. It is certain that a single major geological unit, such as the Hercynian (Carboniferous) chain of Europe or the 300 to 360 million year and 1000 million year metamorphic belts of eastern North America, is formed over a comparatively narrow range of time, but whether apparently coincident events on a world wide scale are cogenetic or fortuitously associated is not yet established. A number of dated mountain ranges are indicated on the map by double lines. The most ancient rocks appear both at the margins and in the centres of continents, but no other generalizations about the distribution of dates are yet possible.

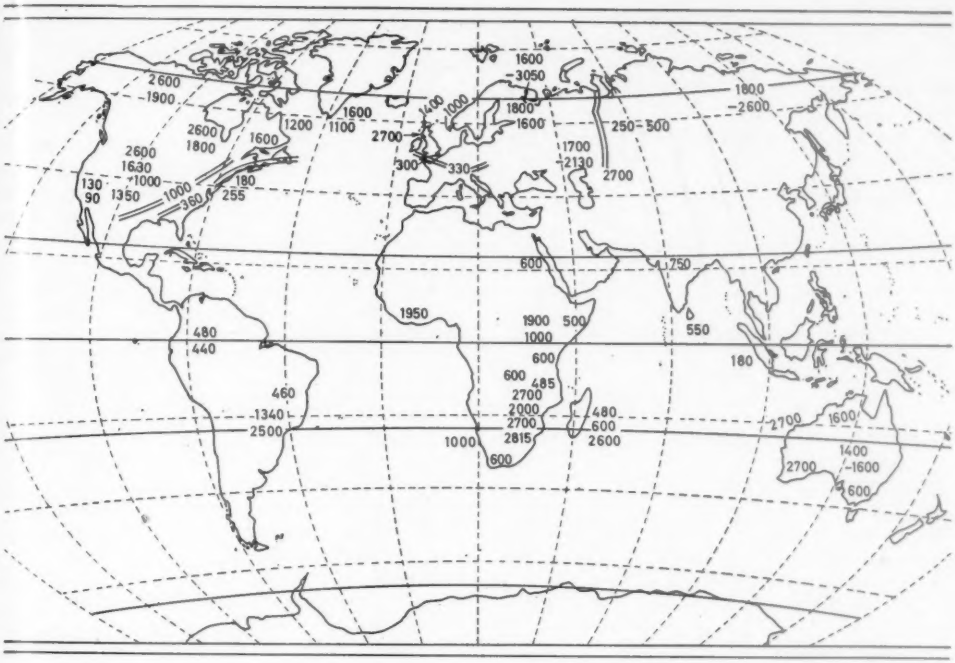


Figure 2. Apparent age of the crust of the earth in millions of years

The oldest known minerals (galena, PbS) come from Southern Rhodesia and are about 3200 million years old. Clearly the earth's crust is at least this old and the search continues for still older rocks. Meteorites have been dated and give an age of 4600 million years. It is generally assumed that this is also the age of the earth and of the solar system. If this is so, there are, as yet, no known rocks which formed during the first third of the earth's history.

At the other end of the geological time scale, fossils become abundant in rocks which are younger than 600 million years and have been used to establish a relative chronology of events for this time. It should now be possible to reverse the process and establish rates of evolution by correlating absolute dating with the appearance of new forms of life. As dating methods become refined, this approach will become increasingly fruitful.

Of considerable interest is the period of time during the last sixty million years. Because it is the most recent, the clearest records are available and the geological record is known in greatest detail. Dating by the long range methods becomes difficult now owing to the minute amount of the daughter nuclide produced. The potassium–argon method shows great promise in this range however. Already, dates have been measured in the thousands of years range and even though they are subject to a large error, they extend into the C^{14} range and permit complete time coverage. It is expected that the potassium–argon method will play a major role in studies on the evolution of man.

References

[1] ARNOLD, J. R. and LIBBY, W. F. *Science* (1949) **110,** 678

[2] ADAMS, J. A. S. and MAECK, W. J. *Anal. Chem.* (1954) **26,** 1635

[3] ALDRICH, L. T. *et al. J. Geophys. Res.* (1956) **61,** 215

[4] TILTON, G. R. *et al. Trans. Amer. geophys. Union* (1957) **38,** 360

[5] COBB, J. C. and KULP, J. L. *Bull. geol. Soc. Amer.* (1957) **68,** 1711

[6] MOORBATH, S. *Nature, Lond.* (1959) **183,** 595

[7] DAVIS, G. L. and ALDRICH, L. T. *Bull. geol. Soc. Amer.* (1953) **64,** 379

[8] ALDRICH, L. T. *et al. Phys. Rev.* (1956) **103,** 1045

[9] TAYLOR, S. R., EMELEUS, C. H. and EXLEY, C. S. *Geochim. Cosmochim. Acta* (1956) **10,** 224

[10] ALDRICH, L. T. and WETHERILL, G. W. *Ann. Rev. nuclear Sci.* (1958) **8,** 257

SOLID STATE DEVICES

I. J. RICHMOND

British Scientific Instrument Research Association, Chislehurst

Following on the earlier article in the series (July, page 254) which dealt with infra-red photocells only, Mr Richmond surveys some of the devices that have been made possible and in particular those using indium antimonide. This is the last article in the series on the technology of semiconductors—the first article was published in September, 1958—and the whole series will be reprinted as a booklet early in 1960.

THE SPEED of technological advance in the field of transistor fabrication and circuitry has been so great during the past decade that the electronic engineer now has the choice of using transistors or valves in his circuits in all applications except for high power transmitters. In future the emphasis will not be on the desirability or otherwise of making a piece of equipment 'fully transistorized', or of ensuring that it contains only 'reliable valves' but on using the best tool for the particular job, so that any particular equipment may well contain valves, transistors, magnetic circuits, and various other electronic devices, each used where it functions most efficiently from the point of view of performance, reliability and cost. This article will be concerned with some of these other electronic devices and in particular with those making use of electronic conduction processes in the solid state. Such devices may well replace valves or transistors in many applications but will have the greatest usefulness where their special properties make them unique. The use of bismuth telluride as a thermo-element and the applications of compound semiconductors to infra-red devices have already been or will be dealt with in other articles in this series and it is not proposed to touch on them further here. Nor will it be possible to deal with such solid state devices as solar cells, parametric amplifiers, masers, nuclear batteries or photoelectric and luminescent cells.

Compound Semiconductors

The demonstration, in 1948, of transistor action in germanium and later in silicon and the successful commercial application of this effect naturally gave rise to a search for other semiconductor materials which could also be used for this purpose and which might serve to overcome some of the limitations apparent with these two substances. Accordingly, many laboratories in various parts of the world embarked on programmes of research into the production and analysis of other possible semiconductor materials. It quickly became apparent that there were no other elemental semiconductors of practical value but there were an enormous number of compound semiconductors possible.

Transistor action can only be obtained with semiconductor material of quite exceptional purity and preferably in single crystal form. The technical difficulties attendant on the production of silicon, one of the commonest naturally occurring elements, in this state has meant that its introduction into large scale transistor production has lagged some years behind that of germanium, even though the latter is naturally very rare. The preparation of the compound semiconductors in comparably pure form has also been attended with various difficulties, not all of which have yet been overcome, but in the case of many of the substances it has been possible to obtain sufficiently pure material to assess their properties and potential uses. The most thoroughly investigated compounds have been those of the 'zinc blende' type in which an element from Group IIIA of the periodic table is combined in stoichiometric proportions with one from group VA. Elements in group IIIA have three valence electrons per atom while those in group VA have five; thus a compound of equal numbers of atoms of an element from each group will have an average of four valence electrons per atom, the same number as silicon and germanium both of which lie in group IV, and will crystallize in the same characteristic tetrahedral or diamond form. Analogous physical properties can be expected as a result of this similar crystal structure.

Properties of Semiconductors

Two of the fundamental properties of semiconductors which are of great interest in assessing their potentialities are the energy gap and the carrier mobility. *Table 1* shows the energy gap and the electron and hole mobilities found for a number of binary compounds together with those of silicon and germanium. Some of the figures given may have to be revised when more accurate measurements are made using purer material.

Table 1

Compound	*Energy gap (in* eV)	*Mobilities (in cm²/volt sec)* ELECTRONS	HOLES
Si	1·1	1200	500
Ge	0·68	3800	1900
AlP	3·0	—	—
Al As	2·2	1200	200
Al Sb	1·6	200	460
Ga P	2·4	1000	20
Ga As	1·35	6000	200
Ga Sb	0·7	4000	850
In P	1·25	4000	650
In As	0·35	30,000	200
In Sb	0·18	80,000	750

Of the various binary compound semiconductors only gallium arsenide and indium phosphide appear to offer the prospect of improved transistor performance as regards frequency and temperature range over that possible with germanium and silicon. Transistor action has been demonstrated with both materials[1], but it may be some time before such transistors are available commercially, though diodes using gallium arsenide which can be used at temperatures up to 325°C are already on the market. However the application of semiconductors to transistors is not their only possible use, though it is obviously the one that has received the most publicity. Some of the other applications can, in fact, be met much better using the compound semiconductors than by using silicon or germanium. In particular the low energy gaps and high electron mobilities of indium antimonide and indium arsenide make possible a whole range of applications in the instrumentation and control fields.

One interesting and useful property of the compound semiconductors is that many of them can be combined in various proportions to give mixed crystals whose properties are intermediate between those of the two constituents. Thus indium arsenide and indium phosphide can be combined in different proportions to give a range of mixed crystals with energy gaps varying from that of pure indium arsenide to pure indium phosphide. In certain applications such mixed crystals may be used to advantage.

Applications of Compound Semiconductors

Very large numbers of the suggested applications of semiconductors make use of galvanomagnetic effects in the material. The Hall effect and the related magnetoresistance effect are two of these. A third, the photoelectromagnetic effect is made use of in infra-red radiation detectors as has already been described in a previous article[2] in this series. In view of its importance the Hall effect will be described in more detail.

The Hall Effect

Figure 1 shows a rectangular piece of semiconductor crystal of length l, breadth b and thickness t. It is situated in a magnetic field H. A current I is passed through the specimen by means of ohmic contacts at the two ends. The direction of current flow is at right angles to the field. Due to the deflection of the charge carriers by the magnetic field a voltage will be developed across the specimen in a direction at right angles to both the current and the magnetic field. This voltage is known as the Hall voltage, after the discoverer of the effect, and if contacts are applied as shown a current may be drawn from these contacts.

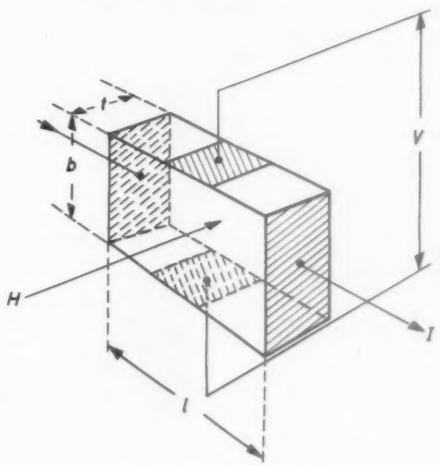


Figure 1. Hall specimen

It should be noted that the effect can be produced in metals as well as semiconductors but in metals the Hall voltage is measured in microvolts whereas in semiconductors with practical values of current and field it may reach several hundred millivolts.

The Hall voltage is given by the equation

$$V = 10^{-8}\frac{BHI}{t} \qquad \ldots.(1)$$

where B is the Hall constant in cm³ per coulomb, H is in gauss and I in ampères. If the Hall electrodes are connected to an external load of resistance R_L then the current flowing in the load is

$$\frac{V}{R_L+R_H}$$

where R_H is the resistance of the specimen measured between the Hall electrodes. The internal resistance R_H is given by

$$R_H = \frac{K\rho b}{Lt}$$

where ρ is the resistivity of the material of the specimen and K is a constant. K takes account of the fact that the Hall electrodes do not extend the full length of the crystal: it is usually found to have a value between two and five.

The maximum power that can be delivered to the load is that given when the load is matched to this internal resistance, *i.e.* when R_L equals R_H.

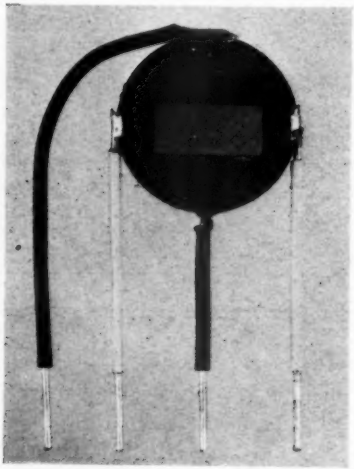


Figure 2. Indium antimonide Hall element (by courtesy of G. V. Planer Ltd)

Under these conditions the power dissipated in the load is

$$W_L = \frac{V^2}{4R_L} = \frac{V^2}{4R_H} = \frac{V^2lt}{4K\rho b}$$

The resistance between the current electrodes is

$$R_I = \frac{\rho l}{bt}$$

so that the power supplied to the specimen W_I is

$$W_I = I^2R_I = \frac{I^2\rho l}{bt}$$

If we define the efficiency of the device E as the ratio of the maximum available output power to

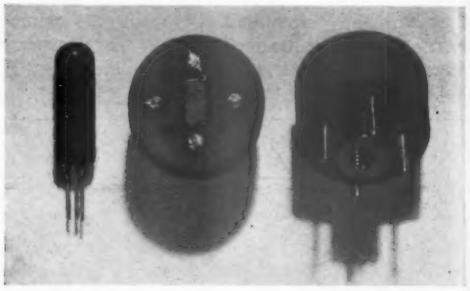


Figure 3. Hall effect unit compared with transistor (after Hilsum; Crown Copyright Reserved)

the input power we have

$$E = \frac{V^2lt}{4K\rho b}\cdot\frac{bt}{I^2\rho l} = \frac{V^2t^2}{4KI^2\rho^2} \qquad \ldots.(2)$$

Substituting the value of V given by *Equation 1* we have

$$E = \frac{10^{-16}B^2H^2}{4K\rho^2} \qquad \ldots.(3)$$

It can be shown that the value of the Hall constant B is given by

$$B = -\frac{3\pi}{8e}\cdot\frac{n\mu_n^2 - p\mu_p^2}{(n\mu_n + p\mu_p)^2}\ \text{cm}^3/\text{coulomb}$$

in this expression e is the electronic charge in coulombs, n and p are the electron and hole concentrations and μ_n and μ_p are the electron and hole mobilities in cm² per volt second. The Hall constant thus depends on the ratio of holes and electrons available as current carriers and on their relative mobilities. With many materials the electron mobility is very much greater than the hole mobility and it is possible to ignore the presence of the holes even in intrinsic material. The equation then becomes

$$B = -\frac{3\pi}{8ne} \qquad \ldots.(4)$$

This equation can also be applied to extrinsic n-type material having comparable hole and electron mobilities provided that $n \gg p$. In both types of material the conductivity is due principally to the movements of electrons and we may write

$$\text{Conductivity} \simeq ne\mu_n = 1/\rho \qquad \ldots.(5)$$

Substituting *Equations 4* and *5* in *2* we have

$$E = 10^{-16} \cdot \frac{9\pi^2}{64n^2e^2} \cdot \frac{H}{4K} \cdot n^2e^2\mu_n^2$$

$$= \frac{9\pi^2}{256K} 10^{-16} H^2\mu_n^2 \quad \ldots.(6)$$

Thus in a given magnetic field the efficiency with which such a Hall specimen transfers power from the current source to the load is proportional to the square of the electron mobility. For instance in this application indium antimonide may have a transfer efficiency of twelve per cent and is some three hundred times more efficient than *n*-type germanium.

It should be pointed out at this point that there are many applications of the Hall effect where the specimen is used in conjunction with an amplifier of very much higher input impedance than the internal resistance R_H, which in the case of intrinsic indium antimonide may well be of the order of an ohm. Under these conditions the Hall terminals are virtually open-circuited and it is the Hall voltage and not the Hall power which will be the most important consideration in determining the relative merits of different materials. *Equation 1* shows that for equal sized specimens and equal fields and currents the Hall voltage will depend directly on the Hall constant.

The Hall constant of intrinsic indium antimonide is 360 cm³/coulomb whereas the Hall constant of *n*-type germanium of five ohm cm resistivity is 21,000 cm³/coulomb. Intrinsic indium antimonide has a high temperature coefficient which may limit its usefulness in some applications. Indium arsenide and the mixed crystal indium arsenide–phosphide are much better in this respect and may replace indium antimonide in many applications.

Hall Effect Devices

Some examples of Hall effect devices are discussed below. A Hall effect device is shown in *Figure 2*; and in *Figure 3* a Hall effect unit is compared with a transistor.

Flux Meter[3,4]

The measurement of magnetic fields has always been more difficult than, say, the measurement of current or voltage. The use of the Hall effect makes possible the construction of simple robust instruments which are capable of measuring directly fields down to a few gauss. There is no need of mechanical motion or variation of the magnetic field. The use of an alternating current supply to the current electrodes of the Hall probe will result in an alternating Hall voltage. Amplification of this alternating voltage will extend the sensitivity of such a flux meter by several orders of magnitude. In extensive weak fields magnetic concentrators may be used. Germanium has been used for this purpose for some years. Indium antimonide with its much greater power output gives very much greater sensitivity. Where stability with temperature is important probes of indium arsenide or one of the mixed crystals have been used. The total thickness of the probe in its protective case can be made less than one millimetre or the semiconductor strip can be mounted between ferromagnetic pole pieces.

Compass

Indium antimonide has been used to measure the earth's magnetic field. Mumetal rods are used to concentrate the field on the specimen. I. M. Ross, E. W. Saker and N. A. C. Thomson[5] describe such a compass (*Figure 4*) in which the attainable sensitivity is such that approximately 10^{-8} watt of output power is produced per degree change around the null position, this being sufficient to give a significant deflection of a sensitive meter.

Clip-on meter[6]

The magnetic field responsible for the Hall voltage in a specimen can be that associated with the current carried by a wire or buss bar. Thus if the Hall specimen is placed in a small air gap of a yoke of high permeability magnetic material, clipped round the conductor, the Hall voltage will give a measure of the current flowing without the necessity of breaking into the conductor. Measurement of currents down to about one milliamp is possible in this way. The device can also be used as a signal tracer in wires and cables. It is applicable to both alternating and direct current.

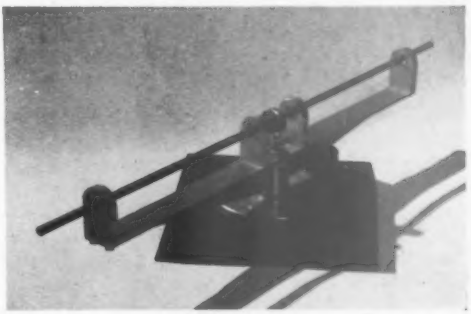


Figure 4. Hall effect compass[5] *(Crown Copyright Reserved)*

Multipliers

The fact that, under certain conditions, the Hall voltage is proportional to the product of the current and the magnetic field makes the device very useful as a multiplying or squaring element in analogue

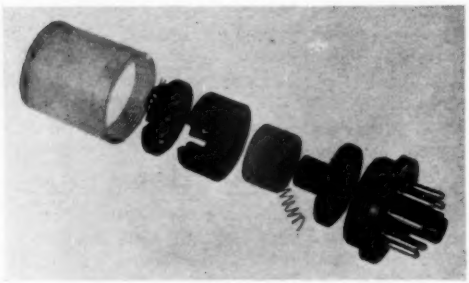


Figure 5. Multiplier Unit (after Chasmar[7]; by courtesy of Metropolitan-Vickers Electrical Co. Ltd)

computers. By using more than one Hall generator three or more quantities can be multiplied. By suitable circuitry reciprocals and quotients of quantities can be obtained. R. P. CHASMAR and E. COHEN[7] have described the precautions necessary to obtain an accuracy of one per cent for an analogue multiplier such as the example shown in *Figure 5*. L. LÖFGREN[8] has reported an accuracy of 0·1 per cent. The device can also be used for comparing the relative phases of two signals of the same frequency.

Modulator

If the control current is at radio frequency and the magnetic field is driven at an audio frequency a modulator is obtained. Suppressed carrier modulators are possible and A. R. BILLINGS and D. J. LLOYD[9] have shown that 80 dB suppression of the carrier can be achieved.

Power Meter[6,10]

A particular application of the multiplier is to the measurement of power. The current through the specimen is made proportional to the current flowing into a load and the magnetic field applied to the specimen is made proportional to the voltage across the load. The Hall voltage gives a measure of the mean power supplied to the load irrespective of waveform or phase angle. H. M. BARLOW[11] has applied the method to the measurement of power at ultra high frequencies. If properly situated in a coaxial line or waveguide the specimen will automatically be subject to the appropriate magnetic field.

Amplifier

An interesting possibility is the use of the Hall multiplier as a chopper for a d.c. amplifier. A typical Hall multiplier[12] mounted in a mumetal magnetic core is rated to give an open circuit output of 200 mV with a control current of 500 mA and a field excitation of 70 ampère turns. Assuming a winding of 70,000 turns a current of 10^{-8} amp in the coil should give a Hall voltage of 2 μV a reasonable figure for applying to the input of an a.c. amplifier. While not of very high sensitivity the Hall chopper is attractive, because of the possibility of using very high chopping frequencies with the consequent gain in overall bandwith. The noise level of such a chopper is close to that of Johnson noise. E. WOLFENDALE[13] has reported a sensitivity of 10^{-7} amp from such a device. (An indium antimonide amplifier is shown in *Figure 6*.)

Motor Torque Measurement and Control[14]

If the Hall generator is mounted on the armature of an electric motor in such a way that it is subjected to the same current flow and magnetic field as a typical conductor, the Hall voltage will give an

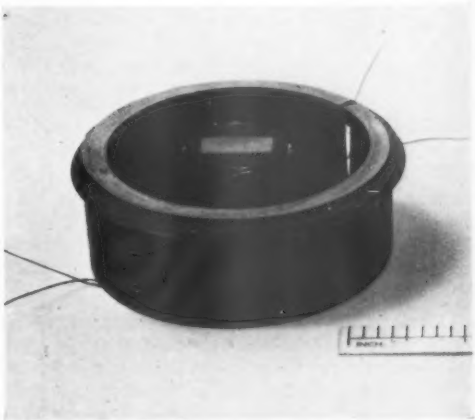


Figure 6. Indium antimonide Hall amplifier with coil assembly removed—this uses very thin sections, 0·0008 in. which together with the attached pole pieces are encapsulated in a high temperature resin (by courtesy of A. V. Planer Ltd)

indication of the instantaneous torque of the motor during actual operation. The voltage can be fed via slip rings for observation on an oscilloscope or it can be used for controlling the torque under conditions of varying load or for operating a relay to give protection against overload.

Susceptibility Bridge

C. Hilsum[15] has described a highly sensitive susceptibility bridge using an indium antimonide Hall element as the detector. A magnetic bridge having four air gaps is unbalanced when materials of magnetic permeability other than unity are inserted in the air gap of one of the arms. The use of a transistor oscillator and amplifier leads to a robust portable instrument suitable for geological surveys. The instrument will measure susceptibilities of the order of 10^{-7} c.g.s. units.

Frequency Spectrum Analyser

V. A. Bogomolov[16] has described a spectrum analyser using the Hall effect in which the signal to be analysed is fed to the current electrodes of the specimen. The current supply to the magnetic field coil is varied in frequency. When the frequency of the magnetic field coincides with a component of the signal being analysed a d.c. output is obtained from the Hall electrodes which can be detected by a galvanometer. Provided the magnetic field strength is constant at all frequencies the output voltage is proportional to the amplitude of the harmonic under investigation. The resolution of the device is very high: so high that some form of automatic frequency control of the variable frequency may prove necessary in a practical instrument.

Gyrator or Non-Reciprocal Four Terminal Network

By suitable arrangements of shunting resistors a non-reciprocal four terminal network can be constructed using a Hall element. Ross and Saker[6] using indium antimonide have obtained a reverse loss of 100 dB with a forward loss of 7·5 dB. W. J. Grubbs[17] has described a three port circulator using germanium. Clockwise and anticlockwise circulation of a.c. or d.c. signals are possible. The measured reverse loss was 61 dB with a forward loss of 17 dB. By optimizing the magnetic field a reduction of the forward loss to half this value is predicted or negative resistances could be used as the shunt elements resulting in zero or negative loss, *i.e.*, gain.

Magnetoresistance

If the Hall voltage produced when a semiconductor specimen carrying a current in the presence of a magnetic field is shorted, either externally or by choice of specimen shape, a current will flow across the specimen between the Hall electrodes. This current will flow in a direction at right angles to the initial current and will itself set up a secondary Hall voltage.

This secondary Hall voltage will be in such a direction as to oppose the initial current flow. This appears to the driving source as a change of resistance of the specimen in the presence of a magnetic field. This magnetoresistance effect can be optimized by shaping the specimen as is shown in *Figure 7*. The relative change of resistance is shown for various shapes in indium antimonide in fields up to 10 kilogauss.

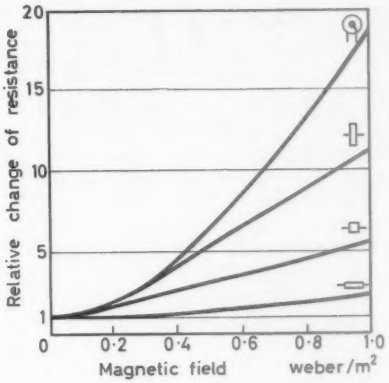


Figure 7. Magnetoresistance effects in indium antimonide

Magnetoresistance Devices

Five magnetoresistive devices will be discussed below.

Variable Resistance or Potentiometer[18]

A variable resistance or potentiometer without moving contacts can be obtained by moving a suitably shaped specimen of indium antimonide or indium arsenide into or out of a permanent magnetic field. Smooth control is possible even for low values of resistance; a feature which may not be easy to obtain by other means. Where a variable magnetic field is employed, a means of control is provided for high current, low voltage systems such as servos, battery chargers and power supplies for transistor circuits.

Relay or Switch[19]

The change of resistance of the semiconductor specimen as it moves into or out of a magnetic field can be used to operate a relay or other switching device. No contact is necessary and quite large powers can be switched in this way. A possible application might be to railway signalling between track and locomotives; on a smaller scale the same effect could be put to use as a limit switch in meters and process controllers.

Displacement Gauge[20]

Any mechanical movement that can be linked to a specimen situated in a non-uniform magnetic field will result in a variation of resistance. The resultant change in current through the specimen can be used as a measure of the movement. (An angstrometer, a device used for measuring small displacements using the magnetoresistance effect in indium antimonide is shown in *Figure 8*.) A bridge circuit is used to reduce the temperature dependence with two specimens arranged to give a differential effect. The sensitivity of such a device is very high. It can be used as a strain gauge, an accelerometer, pressure gauge, thickness gauge or surface roughness meter.

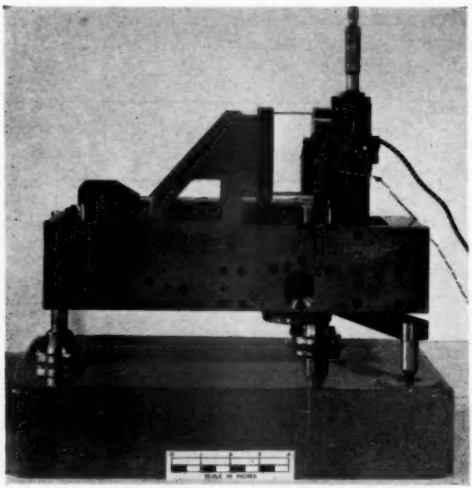


Figure 8. The angstrometer, a device for measuring small displacements using the magnetoresistance effect in InSb (Crown Copyright Reserved)

Microphone or Pick-Up

A special application of the displacement gauge is as a pick-up for a gramophone. The output power from an indium antimonide pick-up is potentially much greater than that of conventional types. The possibility of building a pick-up capable of driving a loud speaker directly has been suggested[19].

Amplifier[18,21]

If the signal current is used to energize the field winding of an electromagnet the resistance of a semiconductor specimen will vary accordingly. The variation of resistance is used to control the current supplied to a load from a battery. The gain per stage is estimated as 30 dB at room temperature using indium antimonide. If the specimen is cooled to liquid nitrogen temperature an increase of gain to 60 dB is possible.

Conclusions

An indication has been given of some of the possible applications of the III–V semiconductor compounds and in particular those of high mobility such as indium antimonide and indium arsenide. The uses of many of the others are only just beginning to be explored as they become available in sufficiently pure form. Beyond this there is the possibility of developing ternary and even more complex compounds which may well find a use in some of the applications described or in other fields not mentioned such as thermoelectric generators and refrigerators, photocells, solar batteries, light amplifiers, parametric amplifiers, masers and of course transistors and diodes. It would seem almost inevitable that it will eventually be possible to select a semiconductor which is virtually tailor-made for any particular application in semiconductor electronics.

The author is grateful to the Council and Director of the British Scientific Instrument Research Association for permission to publish this article.

References

1. JENNY, D. A. *Proc. Inst. radio Engrs* (1958) **46,** 959
2. ROBERTS, D. H. *Research, Lond.* (1959), **12,** 260
3. SAKER, E. W., CUNNELL, F. A. and EDMOND, J. T. *Brit. J. appl. Phys.* (1955) **6,** 217
4. HENNIG, G. R. *Electr. Manufacturing* (1958), **61, 4,** 132
5. ROSS, I. M., SAKER, E. W. and THOMSON, N. A. C. *J. sci. Instrum.* (1957) **34,** 479
6. — — *J. Electronics.* (1955) **1,** 223
7. CHASMAR, R. P. and COHEN, E. *Elec. Engng* (1958) **30,** 661
8. LÖFGREN, L. *J. appl. Phys.* (1958) **29,** 158
9. BILLINGS, A. R. and LLOYD, D. J. *I.E.E. Convention on Transistors and Associated Semiconductor Devices* (1959)
10. STRUTT, M. J. O. *Elect. & Radio Engng* (1959) **36,** 2
11. BARLOW, H. M. *Proc. Inst. radio Engrs* (1958) **46,** 1411
12. Catalogue of Siemans Hall Generators (Munich: Siemens and Halske)
13. WOLFENDALE, E. *I.E.E. Convention on Transistors and Associated Semiconductor Devices* (1959)
14. KUHRT, F. and BRAUENSTREUTHER, E. *Siemans Z.* (1954) **28,** 370
15. HILSUM, C. and ROSE-INNES, A. C. *Nature, Lond.* (1958) **182,** 1082
16. BOGOMOLOV, V. A. *Zhur. Tekh. Fiz.* (1956) **26,** 693
17. GRUBBS, W. J. *Proc. Inst. radio Engrs* (1959) **47,** 528
18. WILLARDSON, R. K. and BEER, A. C. *Electr. Manufacturing* (1956) **57, 1,** 79
19. MOSS, T. S. *Proceedings of the Symposium on the Role of Solid State Phenomena in Electrical Circuits* (1957)
20. ROSS, I. M. and SAKER, E. W. *Nature, Lond.* (1956) **178,** 1196
21. CUNNELL, F. A. and SAKER, E. W. *Progress in Semiconductors* (1957) **2,** 35

First Prize in the Waverley Gold Medal Essay Competition

A STUDY OF COMBUSTION IN SUPERSONIC FLOW

R. A. GROSS

Department of Physics, University of California, Berkeley

This paper describes an exploratory experimental search to learn what phenomena are actually encountered when releasing chemical energy in supersonic flow. Steady, stable, standing plane and oblique detonations were produced for the first time in the laboratory. A new ignition hysteresis phenomenon which has important theoretical consequences was observed. Measured wave properties were compared with theoretical predictions for hydrogen–air and methane–air detonations. This research lays the groundwork for new chemical production methods, new hypersonic propulsion systems and provides a powerful tool in the study of chemical kinetics.

A DETONATION is a wave in which an exothermal chemical reaction takes place and which moves with supersonic velocity with respect to the undetonated (reactant) gas. It is characterized by this supersonic propagation velocity and a large pressure and temperature increase across the wave. A detonation differs from a flame (deflagration) in that a flame moves with subsonic velocity. About one hundred years ago, detonations were brought to the attention of scientists and engineers by accidents in coal mines and laboratories. The phenomenon was discovered concurrently by M. Berthelot[1] and E. Mallard and H. Le Chatelier[2] in 1881. For the following seventy-eight years the literature has carried a steady stream of research reports[3] by chemists, physicists, mathematicians and engineers who became interested in this phenomenon. However, all of this effort was devoted to a detailed understanding of the phenomenon and finding ways and means of avoiding its destructive consequences.

In late 1956 in the Deer Park, N.Y. laboratory of the Fairchild Engine Division, we began a research programme, one of the objectives of which was to attempt, for the first time, to stabilize a detonation in the laboratory. All experience previous to this effort concerned tests of detonations propagating along tubes at about 5000 feet per second. Consequently, these were very short time tests which taxed the ingenuity of the experimentalist. In addition, no useful application of the unusual properties of such waves had previously been found because of their transient nature. If such waves could be made to stand still in the laboratory, they could be studied in detail at leisure and possibly put to use.

We designed a high temperature, steady flow, supersonic research tunnel. This Mach 3 tunnel was put in operation in late 1957 and subsequently demonstrated its ability to produce a clean Mach 3 flow field and to run steadily at inlet stagnation temperatures up to approximately 1950°C (3542°F). The tunnel is of fixed rectangular geometry and the test section flow area is approximately 3 in. × 6 in.

Plane Detonation Waves

During the course of our early thermal ignition experiments with hydrogen and air we learned that a static temperature of about 715°C (1319°F) or greater is required to ignite thermally a flowing mixture of hydrogen and air at a static pressure of about two atmospheres. This condition can be obtained behind a normal shock wave in the supersonic tunnel test section with inlet stagnation properties of about 760°C (1400°F) and 110 psia. We also knew that the initial ignition delay time was less than 30 microseconds.

With this knowledge we set about obtaining a patch of air in the supersonic test section with the required conditions for ignition. This was done by generating (using wedges) an oblique shock from each side wall of sufficient strength to result in a Mach reflected shock pattern in the centre of the flow field. A schlieren photograph of this flow field is shown in *Figure 1*. The circular window height represents 5½ inches. In *Figure 1b*, we describe the approximate properties of this flow field.

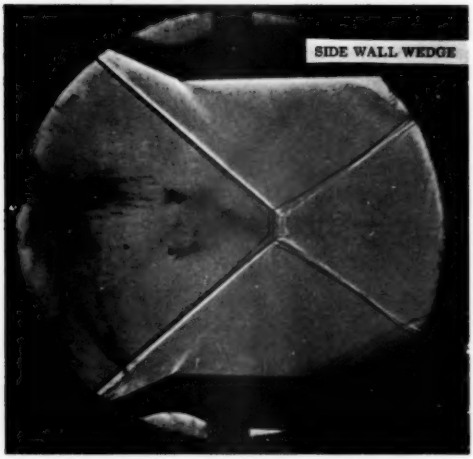
Dr Robert A. Gross is at present a National Science Foundation Senior Post Doctoral Fellow in the Physics Department at the University of California. He was formerly Chief Research Engineer at the Fairchild Engine Division of Fairchild Engine and Airplane Corporation. He is the first United States citizen to have won the Waverley Gold Medal

We then introduced hydrogen from a small orifice located just upstream of the tunnel throat. The hydrogen diffused through the air as it was swept downstream toward the test section. A photograph of the tunnel with its side wall removed is shown in *Figure 2*. A description of some thermodynamic properties along the tunnel centreline for the conditions of a Mach reflected shock in the test section is shown in *Figure 3*.

When hydrogen was added and the tunnel stagnation conditions were brought to conditions (760°C (1400°F), 110 psia) which we hoped would produce ignition behind the Mach reflected wave, the wave shape suddenly changed. The normal section of the wave grew in size and moved upstream to a new stable position. A downstream total temperature probe indicated temperatures considerably above the inlet stagnation tunnel temperature. Changing the amount of hydrogen injected caused a change in the size of the normal wave and its axial location. That is, an increase in fuel flow caused the normal wave to increase in size, move upstream to a new stable location, and the downstream total temperature to rise. A series of schlieren photographs of this new pattern are shown in *Figure 4*.

What are these waves? They appear steady, stable, and amazingly plane in Mach 3 flow for a variety of fuel–air conditions. They could not, therefore, all be the classical, usually encountered Chapman–Jouget waves ($M_2 = 1$*) since there is only one fuel: air ratio that produces such waves at a given Mach number. Since a good portion of the entire wave appeared flat, it seems reasonable to

* Notations used are given in the Appendix



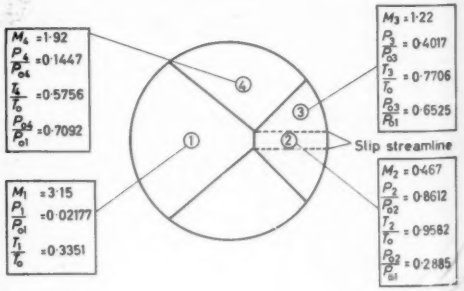


Figure 1. (LEFT) *Schlieren photograph of Mach reflected shock—M_1 is 3·15 and the flow is from left to right;* (RIGHT) *flow field description*

examine the nature of the solutions of the steady, one dimensional flow equations. This was done in detail in a previous publication[4]. Consider *Figure 5* which is taken from that report. Examine the possible supersonic solutions with heat addition. Since, by varying the fuel:air ratio, we essentially vary Q, while the inlet Mach number remains constant, we are either seeing a sequence of strong detonations ($M_2 < 1$) or a series of weak detonations ($M_2 > 1$).

Centreline probe measurements of total pressure and total temperature showed that the flow immediately downstream of the wave was subsonic and corresponded to strong detonations. Schlieren photographs of a probe in this location also showed no shocks, adding further evidence that the downstream flow was subsonic.

Is this simply a shock wave followed downstream by ordinary subsonic combustion? In an attempt to answer this question we did several things. First, what might the flame speed be at these conditions and is this close to the flow velocity behind a Mach 3, normal shock wave? Flame speed data[5] indicates that at 538°C (1000°F), a stoichiometric mixture of

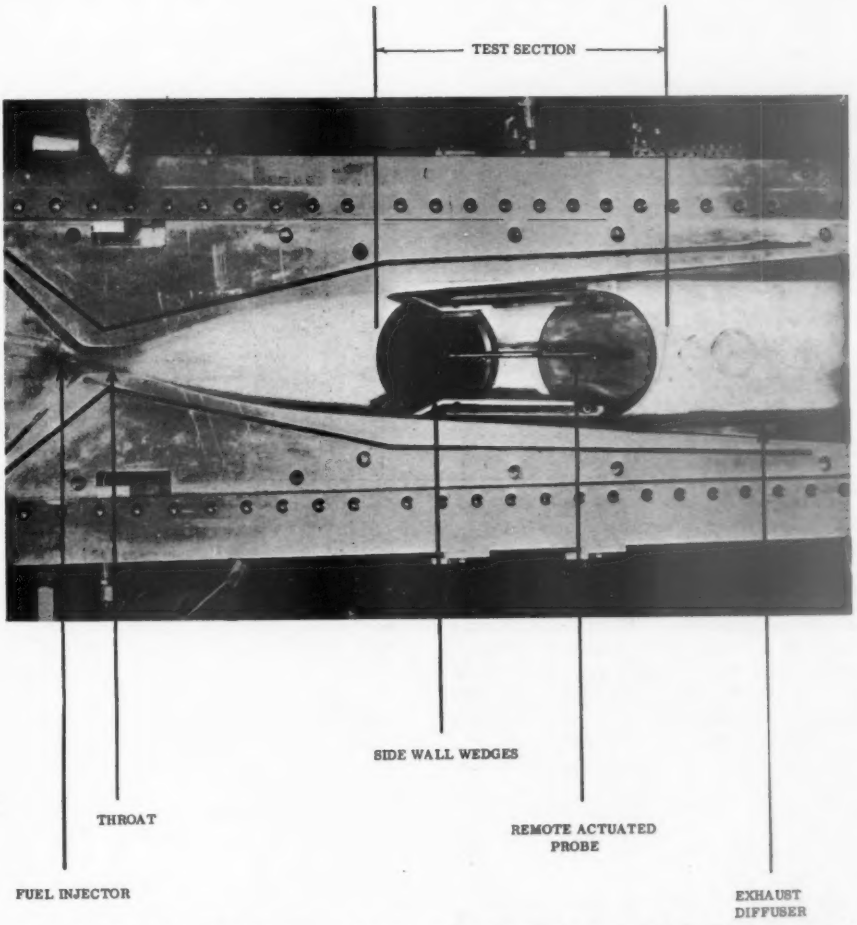


Figure 2. Supersonic combustion research tunnel. The test section is shown with the side wall removed and the wedges used to generate a Mach reflected shock. Note the fuel injector just upstream of the throat and the remote actuated probe along the test section centreline

hydrogen and air has a flame speed of about 40 ft/sec. The velocity behind a Mach 3 shock, $T \approx$ 538°C (1000°F), is about 1000 ft/sec. Consequently, it seems improbable that, even allowing for a considerable margin of error in the flame speed data, a deflagration could be stationary downstream of the shock.

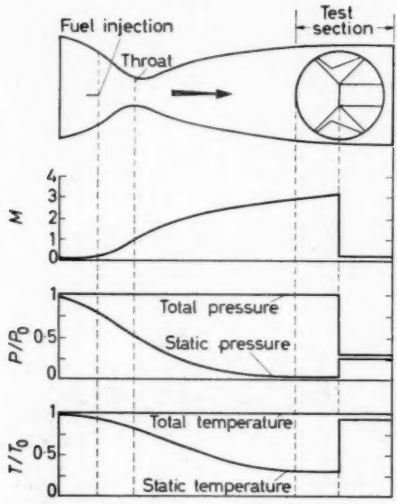


Figure 3. Tunnel centreline properties

Secondly, we decided to determine what effect lowering the inlet stagnation temperature might have upon the characteristics of this wave system. A test was performed in which we raised the inlet stagnation temperature to about 760°C (1400°F), whereupon ignition was obtained and a stable detonation was established. The static temperature behind a normal Mach 3 shock (without combustion) is about 96 per cent of the total absolute temperature. The inlet stagnation temperature was then lowered. To our surprise, the wave maintained itself until an inlet temperature of about 315°C (600°F) was reached, whereupon the wave pattern jumped back to its Mach reflected shock position and probes indicated that combustion had ceased. Our previous experience had indicated that the thermal ignition could not be obtained for any hydrogen–air mixture below about 650°C (1200°F), so this added to the evidence that this was a wave phenomenon and not merely a shock, downstream of which the high temperature mixture caused the fuel to be oxidized.

We conducted tests in which the inlet total temperature was lowered to 260°C (500°F) and the fuel: air ratio decreased so that the total temperature downstream of the detonation wave was about 482°C (900°F). Consequently, nowhere in the entire flow field was there a region of sufficient temperature to thermally ignite the mixture. This hysteresis ignition effect strongly implies that transport mechanisms within the wave itself are important once the combustion process has been initiated. Therefore, the von Neumann–Zeldovich type of analysis which considers the interior of the wave to consist of a pure shock wave followed by chemical reaction in the high temperature shocked gas (with zero transport properties) is too simple. This model does not admit ignition hysteresis phenomenon such as we have found. However, this result does support the recent theoretical work of J. O. HIRSCHFELDER, *et al.*[6,7] in which it is claimed that transport properties within a detonation wave are important.

This ignition 'hysteresis' effect was explored more thoroughly. Again, it was found that the initial ignition of the wave and the subsequent shock to detonation transformation required an inlet stagnation temperature of nearly 760°C (1400°F) and this seemed little affected by the fuel: air ratio. However, once the burning wave was established, the inlet temperature could be reduced greatly, the amount depending upon the fuel: air ratio. Very lean mixtures could withstand only a small temperature decrease while more stoichiometric-like mixtures could go to as low as 204°C (400°F) inlet temperature, the limit of our equipment. This effect was reproducible throughout many tests (in which the test section was cleared of all probes), but the distribution of the points show scatter. The hysteresis effect data are shown in *Figure 6*.

An interesting observation was made during these hysteresis tests. If the fuel: air ratio was kept constant as the inlet stagnation temperature was reduced, the wave increased in size and moved slowly upstream in much the same fashion as caused by an increase in Q. This suggests that the wave location and size is related to the value of

$$\bar{Q} = (Q/C_p T_1),$$

a dimensionless quantity that naturally evolves in the examination of elementary detonation equations.

Finally, a test was performed using sodium as a light emission source to determine where chemical reaction begins in relation to the schlieren wave front location. Some experiments performed by J. A. NICHOLLS *et al.*[8], indicated that a shock wave may be separated from the chemical reaction zone by a distance which corresponds to about 25 microseconds. This possible chemical ignition delay time

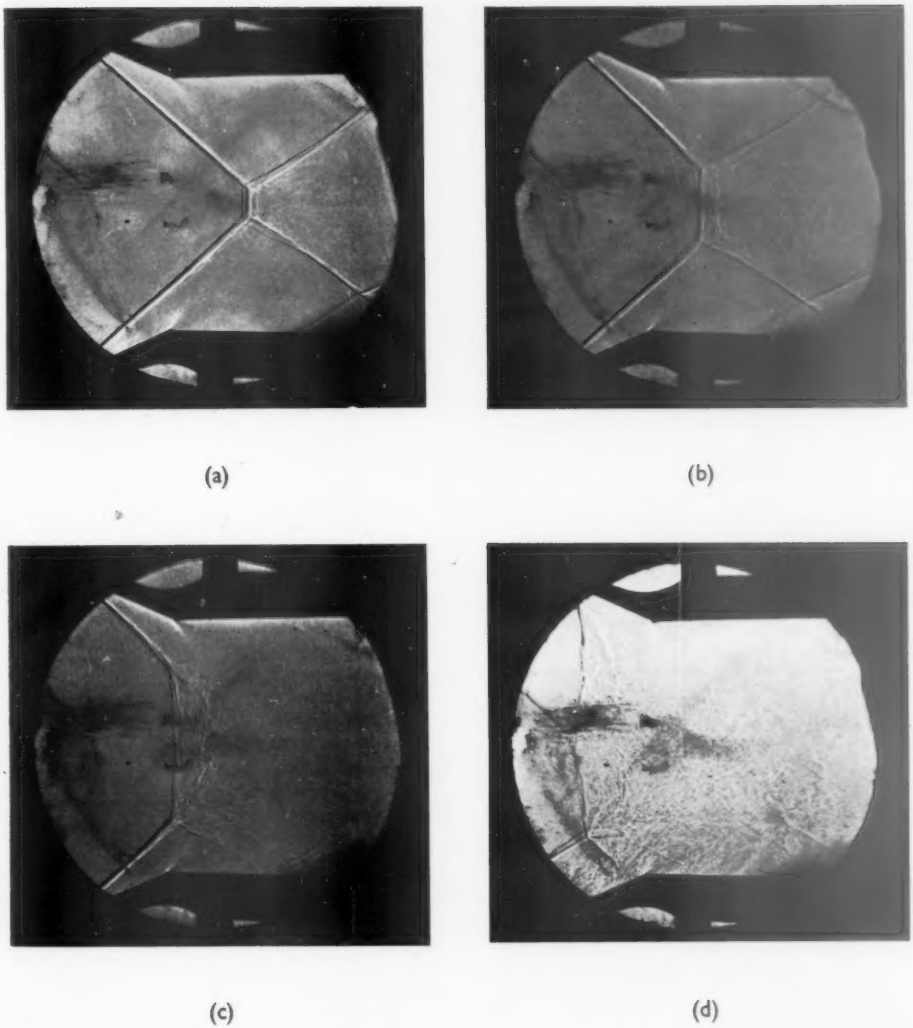


Figure 4. Schlieren photographs of detonation. (a) *Mach reflected wave,* $\overline{Q} = 0$ *(air only)—the flow is from left to right;* (b) *strong detonation wave,* $\overline{Q} = 2{\cdot}75$ *(H_2 and air);* (c) *strong detonation wave,* $\overline{Q} = 3{\cdot}75$ *(H_2 and air); and* (d) *strong detonation wave,* $\overline{Q} = 5{\cdot}0$ *(H_2 and air). The dark smudges upstream of the wave in* (b) *are dirt on the windows*

represents about 0·25 inch, a distance easily discernible in photographs and by eye during the experiment.

A typical schlieren photograph was taken of the burning wave system. Then, common salt (NaCl) was introduced into the plenum air stream by injecting and evaporating a salt–water mixture. The light emitted by the hot sodium atoms was recorded by a film plate in the schlieren optical system, but with the schlieren light source off. A superposition of the negatives from the schlieren photograph and the sodium emission photograph showed that the zone of highest visible radiation intensity occurred immediately behind the wave. These photographs

also indicate that the combustion commences at the wave front. Whatever ignition delay may exist appears to be very small, say less than about five microseconds (corresponding to a distance of 0·05 inch in the photograph).

The foregoing evidence strongly implies that this is an exothermal wave or detonation and not two separate phenomena, namely, subsonic combustion downstream of a normal shock wave.

Is there some chemical reaction taking place upstream of the supersonic test section? Side wall static pressure measurements did not indicate that such was the case. To check this further, and to learn something about the shock to detonation wave transformation, we took high speed schlieren motion pictures of the wave transformation. These photographs, taken at 3000 frames per second, show that the shock to detonation transformation takes place in about 10 milliseconds and in a smooth, continuous manner. There is no evidence that, upon ignition, any part of the wave propagates upstream of the detonation equilibrium position.

To assess the nature of the fuel–air distribution in the test section, as well as to obtain profiles of thermodynamic properties, a remote-actuated traversing probe was designed, fabricated and installed

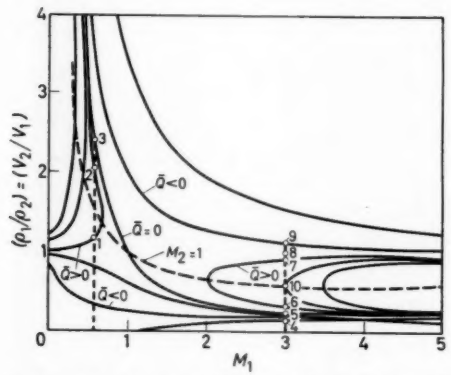


Figure 5. One dimensional flow with heat addition. The numbered points indicate the following: 1 deflagration (simple subsonic combustion); 2 expansion shock + combustion; 3 expansion shock; 4 compression shock + heat withdrawn; 5 compression shock; 6 compression shock + combustion; 7 combustion, no shock; 8 nothing happens; 9 heat withdrawn no shock; and 10 Chapman–Jouget point

in the tunnel. This enabled us to measure temperature and pressure profiles in the vertical and axial direction along the tunnel vertical centreline plane. Samples of gas mixture were also extracted from the tunnel test section and analysed with a mass spectrometer. The vertical fuel–air profile in the test section showed some skewness which is most probably related to the upstream fuel injector

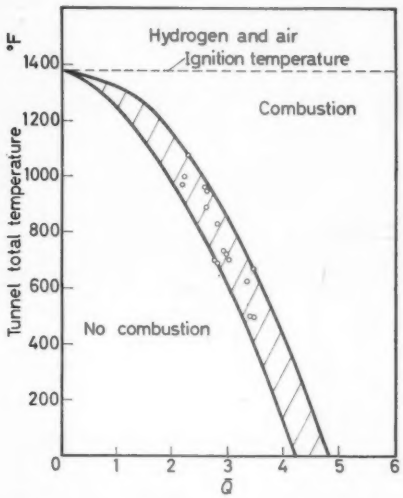


Figure 6. The detonation wave hysteresis effect

alignment. Total temperature and total pressure traverses behind detonation waves for several different fuel flows were obtained. These profiles correspond to the fuel–air distribution, although the location of the triple points of the detonation shock system do not seem to be simply related to the fuel distribution. The total pressure distribution has an M-shape since the traverses extended beyond the triple point and, hence, into the oblique shock wave non-burning region, and then into the boundary layer region.

Since the detonation wave front is flat, how do the downstream properties compare with theoretical, one dimensional, steady, strong, detonation theory? The results of such measurements are shown in *Figures 7* and *8* where total temperature and total pressure measurements are compared with one dimensional theory. The theoretical curves were generated using the digital computer programme described previously[4].

Will fuels other than hydrogen exhibit these same phenomena? Arrangements were made to introduce methane (CH_4) into the tunnel. This fuel was chosen because it contains a carbon atom and has a high vapour pressure at normal temperatures. Consequently, no pumping system is needed to inject gaseous methane into the tunnel. An initial test

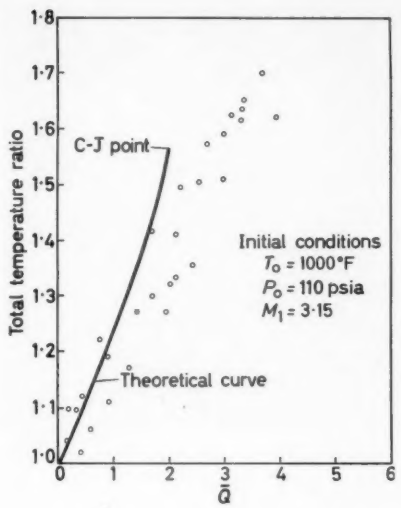


Figure 7. Detonation wave total temperature ratio

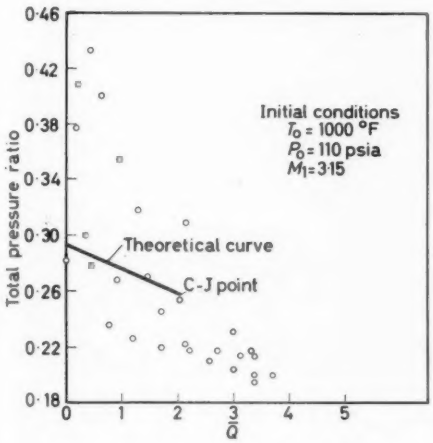


Figure 8. Detonation wave total pressure ratio

failed to ignite the methane. Later tests with nearly pure methane and air produced strong, standing detonation waves of similar configuration to those generated using hydrogen, an initial temperature of over 1090°C (1994°F) was required for ignition. Once started these methane–air detonations also exhibited the aforementioned ignition hysteresis effect.

Oblique Detonation Waves

Can oblique steady detonation waves be obtained in the laboratory? The possibility of steady oblique detonations has been postulated by G. D. SAMARAS[9] and others. The predicted properties of such waves have been extensively treated by W. CHINITZ, L. C. BOHRER and K. M. FOREMAN[10]. Basically, one finds solutions like the classical shock polar; but a family of solutions now exist because of the additional parameter of heat addition. In *Figure 9* (taken from reference 10) are shown the possible solutions for an oblique detonation wave in a Mach 3 flow. Here, δ is the streamline deflection angle, σ is the wave angle with respect to the incoming flow and the dotted line indicates the locus of Chapman–Jouget points; *i.e.*, where the downstream normal velocity component is just sonic.

An experiment was arranged in which a single, 30 degree wedge was placed in the test section. The shock wave angle is about 52 degrees and the downstream Mach number is about 1·4. From *Figure 9*, upon heat addition we expect an increase in the wave angle σ to about 65 degrees at which condition analysis indicates further heat addition will cause the wave angle to decrease. Thus, for a fixed inlet Mach number and deflection angle δ, there exists either two solutions, one solution

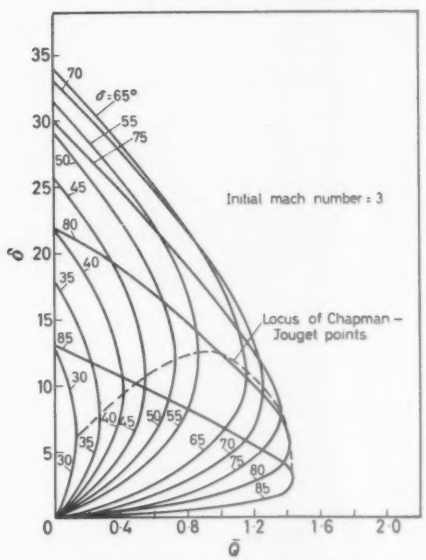


Figure 9. Oblique detonation waves—initial Mach number is three

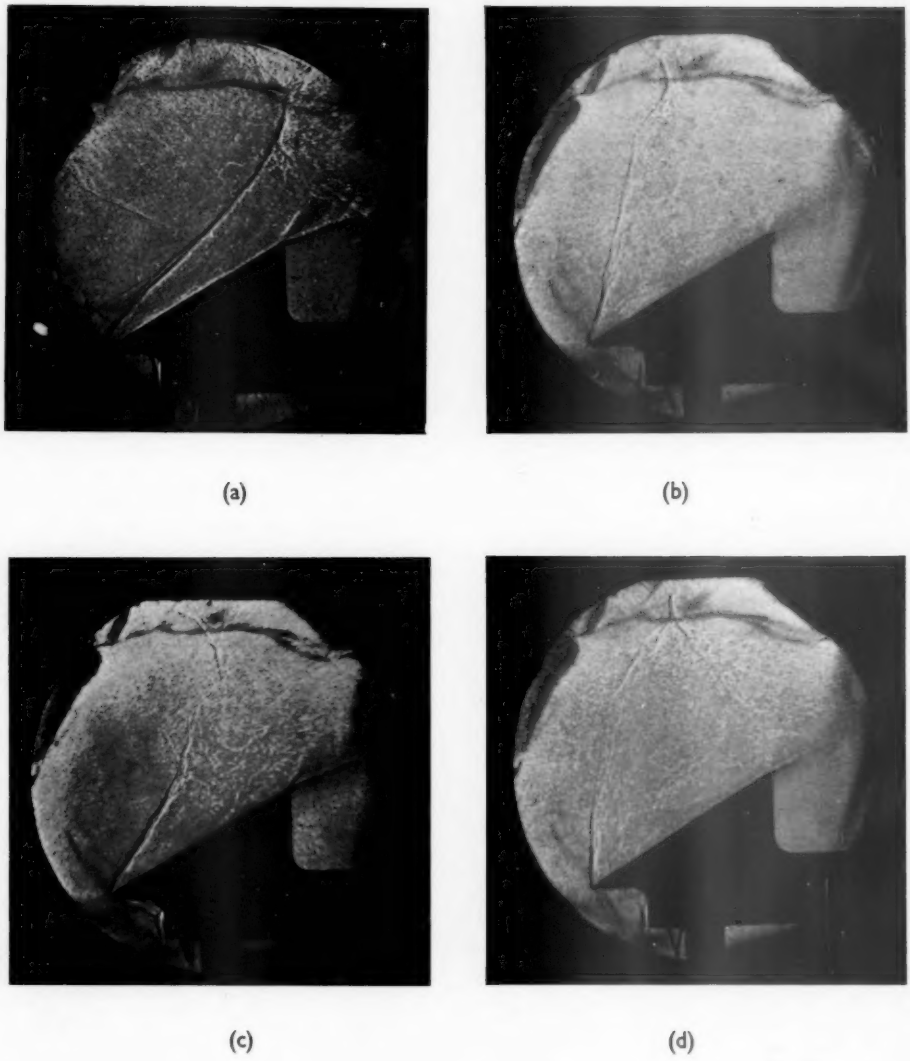


Figure 10. Schlieren photographs of oblique waves. (a) *Oblique shock, $M \approx 3$ and $\bar{Q}=0$—note the weak upstream oblique shocks indicating that the approach flow is not uniform;* (b) *oblique detonation, $M \approx 3$ and $\bar{Q}=0{\cdot}8$;* (c) *oblique detonation, $M_1 \approx 3$ and $\bar{Q} \approx 1{\cdot}3$; and* (d) *oblique detonators, $M_1 \approx 3$ and $\bar{Q} \approx 3{\cdot}1$.*

(maximum $\bar{Q}$), or no steady solutions (greater than maximum $\bar{Q}$).

In *Figure 10* is shown a schlieren photograph of the 30 degree wedge in the supersonic test section at Mach 3·15, with no fuel introduced. Some hydrogen was then injected upstream of the throat and the tunnel stagnation temperature raised to about 1340°C (2450°F), whereupon the wave angle increased. Changing the fuel flow changed the wave angle. However, the behaviour was single valued; namely, an increase of fuel always increased the wave angle. Wave angles up to about 80 degrees were

obtained. In *Figure 10* are shown representative schlieren photographs of these exothermal oblique waves. These oblique waves also exhibited the ignition hysteresis effect and can be kept burning with a low inlet stagnation temperature, about 315°C (600°F). The wave angle increases for a fixed fuel flow as the inlet temperature is reduced, again indicating the importance of the heat addition parameter $\bar{Q}$.

Application

To our knowledge, these were the first stable, standing, normal and oblique detonations produced in the laboratory. The demonstration of the existence of these waves opens up interesting applications of these phenomena. The wide range of actual operating conditions permits engineering use without close control.

Hypersonic airbreathing propulsion systems employing this form of energy release are currently being studied[11]. Such a detonation wave ramjet would have certain advantages such as short length, light weight, and relatively good specific fuel consumption at hypersonic flight speeds. Because of the ignition hysteresis effect, the wave may be stabilized within the engine over a very wide range of flight speeds, altitudes, and fuel flow rates.

The possibility of new chemical production methods is suggested. Mixing of reactive chemicals in gaseous form at supersonic speeds (when the static temperature is low) and then permitting the reaction to occur via a standing detonation at precisely the time and position desired is possible. Substances normally reactive at room temperature can be mixed safely at conditions corresponding to high Mach number flow. At low pressures the detonation reaction zone is spread out in space and intermediary species, such as acetylene, may be extracted. Chemical process plants have not been built for fear of detonation and consequent destruction of vast and expensive equipment. Controlled detonation permits the new design of such plants.

Finally, these phenomena should prove a powerful tool in the fundamental study of chemical kinetics. Basically, a chemical system is exposed to a step function in pressure and temperature. A major molecular change in the mixture results from this change in thermodynamic conditions. All of this takes place continuously and at a fixed place in space. Consequently, absorption and emission spectrographic techniques can be applied to search for the detailed molecular changes, chain making, chain breaking steps, *etc.* Preliminary spectroscopic work already has begun at the Deer Park laboratory and early results are encouraging.

The unusual properties of the phenomena of standing detonation challenge the imagination of engineers and scientists.

This research is actually the work of many persons and some who were more directly concerned with the results presented here are: W. Chinitz, A. Cozewith and A. Farina. The work of Nicholls, et al.[8] *may well have been the first attainment of a standing detonation and his excellent work, done during the same general time period as that reported here, is acknowledged. This paper and the associated research were made possible by the financial support of the United States Air Force office of Scientific Research. It is a pleasure to acknowledge the encouragement and support given by Lt Col. Paul G. Atkinson, Dr Carl Kaplan, and Dr M. Slawsky of the AFOSR.*

References

[1] BERTHELOT, M. *C.R. Acad. Sci. Paris* (1881) p. 18
[2] MALLARD, E. and LE CHATELIER, H. *ibid.* (July 18, 1881), p. 145
[3] GROSS, R. A. and OPPENHEIM, A. K. *Amer. Rocket Soc. J.* (March 1959) p. 173
[4] EISEN, C. L., GROSS, R. A. and RIVLIN, T. J. 'Theoretical Calculations in Gaseous Detonation' AFOSR TN 58–326, ASTIA No. AD 154, 230. To be published in *Combustion and Flame*
[5] DRELL, I. L. and BELLES, F. E. 'Survey of Hydrogen Combustion Properties' NACA RME57D24C (July 26, 1957); declassified Dec. 13, 1957 (data taken from *Figure 8*, p. 62)
[6] HIRSCHFELDER, J. O. and CURTISS, C. F. *J. chem. Phys.* (1958) **28**, 1130
[7] LINDER, B., CURTISS, C. F. and HIRSCHFELDER, J. O. *ibid.* p. 1147
[8] NICHOLLS, J. A., DABORA, E. K. and GEALER, R. L. 'Studies in Connection with Stabilized Gaseous Detonation Waves' Seventh Symposium on Combustion (Aug–Sept. 1958), p. 144 London: Butterworths Scientific Publications, 1959
[9] SAMARAS, D. G. *Canad. J. Research* (Jan. 1948) p. 1
[10] CHINITZ, W., BOHRER, L. C. and FOREMAN, K. M. 'Properties of Oblique Detonation Waves' AFOSR TN 59–462, ASTIA No. AD 215, 267 (April 1959)
[11] SARGENT, W. 'A Detonation Wave Ramjet' AFOSR TN 59–589, ASTIA AD 216, 811 (June 1959)

Appendix

The following notations were used:

M = Mach number
P = Pressure
Q = Heat added to the stream
$\bar{Q}$ = Dimensionless heat addition = Q/C_pT_1
T = Temperature
C_p = Constant pressure specific heat
σ = Wave angle
δ = Streamline deflection or wedge angle

and the subscripts mean:

0 = Total quantity (pressure, temperature, *etc*)
1 = Upstream of wave
2 = Downstream of wave

Special Prize in the Waverley Gold Medal Essay Competition

ENGINEERING DESIGN BY DIGITAL COMPUTERS

B. J. CHALMERS

Faculty of Engineering, The University of Bristol

The author gives a brief description of the fundamental properties of a digital computer, and then indicates how these simple properties may be utilized in performing complex calculations and simulating logical decisions. A few selected examples, extracted from computer programmes for the design of electrical plant, are given in order to illustrate the techniques employed with computers, and the principles are then extended to show how complex design procedures may be arranged for automatic calculation. A general assessment of the importance of this development is presented, and the essay concludes with a summary of progress to date. The essay was awarded the special prize presented for the best entry in the *Waverley Gold Medal Essay Competition* from an author under thirty years of age.

THE RAPID developments in computer design, which took place mainly over a period of about ten years and led to the Convention on Digital Computer Techniques presented by the Institution of Electrical Engineers in April 1956, have been followed by a correspondingly widespread development in the utilization of these computers in an ever widening field of applications. In clerical work and data processing, in production control, scientific calculations and engineering design, computers have been increasingly applied to release man-power from routine and often tedious labour. Due emphasis was given to the extent and importance of these types of applications in the 1958–59 Faraday Lecture on *Automation*, the main theme of which was the use of computers and automatic devices—leading to the displacement of some of man's mental effort by machines.

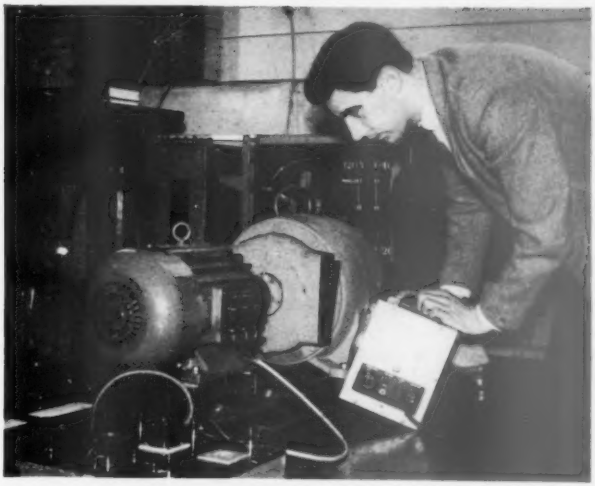
It is the object of this essay to describe some of the principles developed in research into the application of computers to engineering design, with particular reference to the design of electrical plant. For this purpose, it has been deemed advisable to give a general account of the properties of computers and the way in which these may be utilized to perform the work of a designer, rather than to describe a particular computer programme for the design of a single type of equipment. The type of application will be illustrated by a few selected examples, of a comparatively simple nature, which have been extracted from work with which the author has recently been associated.

Use of the Basic Properties of the Computer

In order to reduce the problems to their most fundamental level, all reference to computer technology will be avoided, and the computer itself described simply as a device having storage locations and high speed mathematical facilities. The essay is not concerned with the nature of these, but with their use and application.

The storage positions in a computer are used to store numbers—which will include the data given in the problem, and intermediate and final calculated figures—and also to store the coded instructions which are written to direct the solution of the particular problem. These instructions, which constitute the computer programme, are very often as simple as ' take the number in storage position 5, add it to the number in position 11, and retain the result in position 27 '.

The basic arithmetic facilities of the computer enable addition and subtraction to be performed, and there is also a very simple logical or ' choosing ' function, the nature of which is described later. It is by combination of these simple facilities that the computer experts have built up extensive libraries of programme elements, generally called ' sub-routines ', which perform all the mathematical operations normally required. There are, for example, sub-routines for multiplication, for division, for taking square roots, for curve reading, and for rounding-off to whole numbers. The task of



Mr Brian Chalmers is at present engaged in post-graduate research at the Faculty of Engineering, University of Bristol

the designer-turned-programmer is to utilize the existing facilities and ready-made programme elements in order to meet his own requirements. In writing a programme for a.c. circuit calculations, for example, his labour is greatly reduced by use of the sub-routines for squaring, taking square roots, and obtaining the sines and cosines of angles.

Since the calculations and decisions of a plant designer are essentially of a logical nature, it will be useful to state at the outset that a digital computer has two overall properties which make it potentially useful to the designer. These are:

(*i*) Its ability to carry out long and complex calculations very quickly and, of course, with high accuracy.

(*ii*) Its ability to take a course of action dependent on the answer obtained at a particular stage of the calculation.

The second property, which utilizes the logical facility of the computer commonly called ' discrimination ', may lead to one or other of two alternative courses, depending whether a particular result is positive or negative. This means that if the computer is instructed to form the difference of two significant quantities in a design calculation and then to examine the mathematical sign of the result, the two alternative courses which then arise will lead to two quite different sections of the programme. That is, a logical decision may be built into a computer programme in this manner.

Thus, in order to programme design decisions, it is, in general, the aim to cast the basis of the decision into the form of a comparison of a given quantity either with limits on its value or with some particular desired value. The discrimination facility may then be used to effect the comparison, and the subsequent design action, which is the next section of the programme, depends on the result of the discrimination.

A very simple example of this technique, taken from electrical machine design, is the comparison of the electrical current-density in a conductor with an upper limit of current-density, which is the maximum value that will be tolerated. The procedure adopted for this comparison may then be detailed as listed below and as shown in *Figure 2*:

(*i*) Calculate current-density in the conductor;

(*ii*) Form the difference between the upper limit and the current-density;

(*iii*) See whether the answer is positive or negative; *i.e.*, discriminate on the mathematical sign of the result;

(*iv*) There result two alternative courses of action. A positive answer signifies a satisfactory result, and leads on to the next calculation in the programme. A negative answer indicates that the current-density is too high, and leads to a part of the programme which directs an appropriate design action, such as the selection of larger slot and conductor sizes for the machine.

This simple example typifies the manner in which the thoughts and actions of a designer may be mechanized, in a precise mathematical form, for automatic calculation within a computer programme. A general discussion of the problems which arise in the organization of complex design procedures for the design of complete equipments or machines will next be presented.

The Organization and Programming of Complete Design Procedures

The basic approach used in many existing computer programmes for the design of electrical plant is to obtain first a provisional design based on a set of initial estimations; certain significant characteristics are then computed and compared with the values required by the specification. If the results are unsatisfactory, that is, not within prescribed tolerances, the computer follows the course of action which directs it to adjust the initial estimations suitably, and the design is then recalculated. This procedure is continued until the design is accepted as satisfactory. A simple block diagram illustrating the type of procedure is given in *Figure 3*.

It is evident, therefore, that the main problems to be considered in the construction of a programme for a design of this nature concern the relationships between the various dependent and independent variables. The laws by which the adjustable quantities are modified in each iterative loop determine whether or not the design will converge to meet the given specification. With this in mind, it is obvious that the use of computers in the field considered cannot proceed independently of design and analytical knowledge; indeed, its success depends almost entirely on the experience of designers and the logical formulation of that experience.

It is not within the scope of this essay to consider the design of any particular type of machine in detail, but a reference to one example of some complexity will assist in demonstrating the nature of the general problem.

In the design of a large squirrel-cage induction motor there are considerable numbers of performance figures to be controlled and design variables to be adjusted. Thus, the following quantities have usually to be obtained within specified limits: starting torque, starting current, maximum torque, full-load efficiency, full-load power factor, full-load temperature rise and the temperature of parts during run-up. The design quantities which may be adjusted in order to obtain the desired characteristic are: core length, diameter, number of slots in stator and rotor, number of effective stator conductors per slot, stator slot width and depth, rotor bar material and the end-ring area and material.

Each of these design quantities has its influence on the circuit parameters, *i.e.*, resistances and reactances, which, in turn, determine the performance

Figure 1. An English Electric Deuce computer which was recently installed at the Central Electricity Generating Board, Bankside House, London (by courtesy of The English Electric Company Limited)

of the machine. Without proceeding any further with the problem, it is manifest that the formulation of a logical design procedure for automatic calculation must be firmly grounded on the theoretical

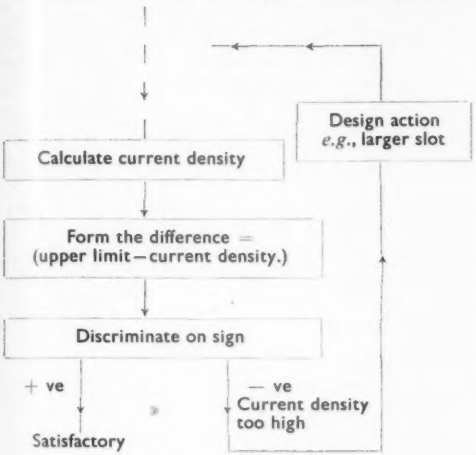


Figure 2. A simple design decision based on a current density

relationships between the two sets of quantities, backed up by design experience. The greater the number of interdependent variables, the more complex must be the design procedure adopted, since a greater variety of possible conditions have then to be taken into account.

It is not surprising that difficulties have arisen in the development stages of the work, since the arrangement of the loops of design adjustments to be followed by the computer *must* ensure that convergence to a satisfactory design will always be obtained. A particular type of danger is that interminable loops of adjustments, or oscillations, may occur if, for example, one stage of an iterative programme requires an increase in a particular quantity and a later stage requires a decrease in the same quantity.

The most illustrative example of this type in the writer's experience occurred in the development of a section of a transformer design programme. The particular section was so constructed as to obtain a temperature gradient which converged to a certain desired value. When a set of results was examined, it was found that the temperature gradient was exactly as required but, alas, the transformer was half an inch tall and three miles long. Some of the manufacturing requirements had obviously been omitted, and this section of programme was hastily revised.

In general, it may be said that this type of fault arises solely due to human error, in the incomplete statement of the problem. The importance of the preliminary stages of investigation, formulation, and analysis, prior to the actual programming of a problem, cannot therefore be over-emphasized: following the construction of an initial programme, successive stages of testing, proving, and improving are, in most cases, advantageous and usually essential.

Development of Techniques for Design Optimization

Having shown, in broad principle, how it is possible to arrange design calculations and decisions in a form suitable for automatic calculation by a digital computer, further advantageous extensions of the technique may now be developed. The high speed of calculation and the logical facility offered by a computer make it possible to improve designs in a manner which is not feasible in design by human effort. In particular, the possibility is introduced of producing an optimum rather than a 'first satisfactory' design, since alternative designs may be compared using the discrimination facility.

The following example of the optimization of a particular section of the design of an electrical machine has recently been suggested[1]. The selection of a stator slot size could be optimized by directing the computer to examine the variation of the total stator power loss—copper loss + iron loss in core + iron loss in teeth + eddy current loss—with variation of slot width and depth and, by an organized process of comparison, thus to select the precise slot size which gives the minimum total loss.

Complete designs may be optimized in a similar manner. In the case of the design of an induction motor to meet a given specification, several alternative designs may be obtained, each of which is satisfactory, but with different diameters and core lengths or different numbers of stator and rotor slots. Since a completed design may be obtained in about three minutes using a high speed computer, it is quite practicable that the computer should compare these designs and select the best one of them.

It is to be emphasized, however, that this type of application necessarily involves an appreciation of the quantitative criterion which will identify the best design for a given duty. In order to choose between a set of alternative designs, the deciding criterion may well be the result of the weighted combination of several factors, such as efficiency, power factor, and capital cost.

In this comparison of designs, economic considerations must play a very important part, and much fundamental study remains to be done before this aspect of design by digital computer can be deemed

to have proceeded beyond its infancy. There is, however, vast scope for the immediate application of optimization techniques to details and sub-sections of designs.

Effects of the Use of Computers in Design Offices

Following the development and application of the first computer programmes for the production of detailed engineering designs, it is possible to summarize the effects of this use of computers and also to estimate further far-reaching consequences.

Results already obtained have shown that the use of computers enables improved designs to be obtained far more speedily than hitherto. The advantages of improved performance are sufficiently obvious to require no further explanation. The new speed acquired is of great importance in dealing with routine sales enquiries, where at present a considerable number of man-hours are spent dealing with designs of which, inevitably, only a proportion lead to actual orders. The growing difficulty in obtaining suitable men to train as designers increases the importance of the saving in time which results. It is also to be hoped that the introduction of computers will release designers from the somewhat tedious routine design work to give more attention to the development problems associated with their products.

Furthermore, there is now a stimulus to the use of more accurate and more lengthy relationships in design calculations, where empirical or approximate relationships have previously been made to suffice. Unlike the human calculator, a computer can perform these lengthy calculations repetitively without experiencing fatigue. The computer may, of course, also be used as an aid in the determination and proving of new relationships, by checking against test figures the results obtained using tentative new expressions.

Conclusions

This essay has demonstrated the types of function a digital computer can fulfil in assisting the engineering designer. Results already obtained have shown that the high speed of operation and the logical facility of digital computers make their application advantageous in both complex calculations and in the automaic calculation of complete designs. Further developments in the technique should lead to particularly outstanding improvement in design office methods and in the results obtained.

An indicative conclusion is provided by a brief summary of some of the achievements reported during the last three years. Computer programmes have been completed which produce designs and performance data for induction motors, power transformers, rural type transformers, current

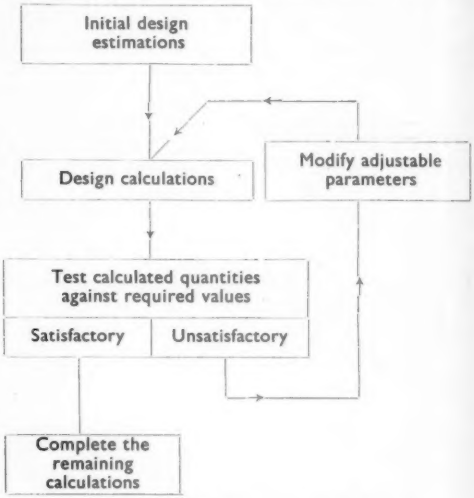


Figure 3. Simple block diagram of design procedure

transformers, certain types of d.c. machines, and many aspects of turbo-alternator designs, in times ranging between one minute and ten minutes. At the same time, similar work has been proceeding apace in the fields of mechanical engineering and aircraft design.

Finally, it is observed that the ultimate test of a programme's validity must always be whether established designers of great experience, responsible for the performance of their products, are prepared to accept the results produced by a machine as being equivalent, and sometimes even superior, to the results of their own mental efforts.

Reference

[1] REECE, A. B. J. and CHALMERS, B. J. Symposium on the Application of Digital Computers to the Design of Electrical Plant (Queen Mary College, London) April 1958

THE GROWTH OF OXIDE SINGLE CRYSTALS CONTAINING TRANSITION METAL IONS

F. W. HARRISON

Mullard Research Laboratories, Salfords, Nr. Redhill, Surrey

Measurement of the fundamental electrical and magnetic properties of oxides containing transition metal ions requires the growth of single crystals. Problems of preparation include high melting point and reactivity, and dissociation with evolution of oxygen below the melting point. Methods are described which overcome some or all of these difficulties, the more important being the flame fusion process, the Bridgman–Stockbarger and the lead oxide solution methods.

DURING the past ten years, due primarily to their technical potentialities, there has been increasing interest in the magnetic and electric properties of oxides containing transition metal ions. In the research field the measurement of basic parameters dependent upon orientation of the crystal structure, or of phenomena sensitive to grain structure has led to a need for single crystals of the materials.

As examples of the materials and their phenomena, we may mention antiferromagnetism and conductivity in MnO, CoO and NiO; ferrimagnetism and the soft magnetic properties of the ferrites $M^{2+}Fe_2^{3+}O_4$, where M can be Mg, Mn, Fe^{2+}, Co, Ni, Cu, Zn, or a combination of these ions; the barium-containing compounds such as the permanent magnet material *Ferroxdur*[1], $BaFe_{12}O_{19}$ and the high frequency *Ferroxplana* materials[2], $BaM_2Fe_{16}O_{27}$, $Ba_2M_2Fe_{12}O_{22}$ and $Ba_3M_2Fe_{24}O_{41}$ where M is a divalent ion such as those listed above; and the magnetic and optical properties of the garnets $M_3Fe_5O_{12}$, where M is Y or a rare earth ion.

The growth of single crystals of such materials presents a number of problems. The availability of phase data up to liquidus temperatures is of importance in determining the method and growth conditions most likely to be successful in a particular case. Whilst the absence of phase data for a given system does not necessarily mean that it is impossible to grow a composition within that system, the data are of value in assessing the chemical condition of the crystals grown. In general, in order to produce single crystals of reasonable size, it is necessary first to melt the material. Since the melting points of the compositions listed above are all greater than 1500°C a high temperature furnace is required, together with a crucible capable not only of withstanding the temperature but also insensitive to attack by the molten oxides. Furthermore, since oxygen is a component of the system in which crystallization is taking place, it is necessary to adjust the oxygen pressure above the melt to the correct value for equilibrium if a crystal with stoichiometric composition is to be obtained. Compositions containing Fe_2O_3 as a constituent tend to lose oxygen rapidly at atmospheric pressure when the temperature rises above 1400°C so that the equilibrium oxygen pressure at the melting point may be considerably more than one atmosphere.

The growth of compositions which melt incongruently presents additional problems, whilst in general it is not possible to grow single crystals of materials which exhibit a phase change below the melting point.

The methods to be described overcome some or all of these difficulties and may be classified under the following headings: Verneuil or flame fusion method, Bridgman–Stockbarger method, flux method, grain growth and others.

Flame Fusion Method

The flame fusion method was invented over fifty years ago by Verneuil[3] for the production of synthetic sapphire and was used extensively during the war, particularly in Germany, for the preparation of jewel bearings[4]. Its most recent scientific application is in the preparation of artificial rubies for use in Maser devices.

The application of the method to the growth of single crystals of rutile has been described by C. H. MOORE[5], whilst W. H. BAUER and his

colleagues[6] at Rutgers School of Ceramics, New Brunswick, N.J., have produced single crystals of mullite, $3Al_2O_32SiO_2$ and a number of ferrites[7–10]. E. J. SCOTT[11] has grown NiO, CoO, Cr_2O_3, $NiFe_2O_4$, $CoFe_2O_4$, $MnFe_2O_4$, $NiAl_2O_4$, $CoAl_2O_4$ and $MnAl_2O_4$ and has claimed Mn_3O_4, Fe_3O_4 and $Fe_{2\cdot25}Al_{0\cdot75}O_4$. In Russia, A. A. POPOVA[12] has prepared Mn, Co, Ni, Zn, Mg–Mn, Mn–Zn, Ni–Zn ferrites. In our laboratory, J. L. PAGE has grown NiO, $MnFe_2O_4$, $Mn_{1\pm x}Fe_{2\mp x}O_4$, $Co_xMn_{1-x}Fe_2O_4$, $Cu_xMn_{1-x}Fe_2O_4$ and has investigated the growth of MnO, $Fe_{3-x}Al_xO_4$ yttrium iron garnet and $BaFe_{12}O_{19}$. More recently, R. G. RUDNESS and R. W. KEBLER[13] have grown successfully single crystals of yttrium iron garnet.

The essentials of the process are as follows. The high temperature necessary to melt the material is obtained by means of an oxy-coal-gas or oxy-hydrogen burner, the flame from which is introduced vertically into a firebrick furnace surround (see *Figure 1*). The burner of the apparatus can be

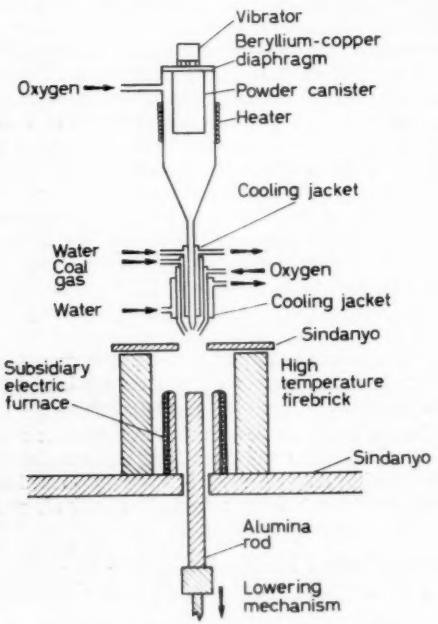


Figure 1. Schematic diagram of the flame fusion apparatus

composed of two concentric tubes, the inner supplying oxygen and the outer coal-gas or hydrogen to the furnace. Alternatively, a third outer concentric tube may be added supplying oxygen to produce a more oxidizing flame whilst providing an additional means of varying the flame conditions. The burner may be made from telescopic brass tubing if adequate water cooling jackets are added to the design. Such cooling can be largely dispensed with if higher melting point steels are used.

The diameters of the burner nozzles are not critical, except that the inner oxygen nozzle must not be so narrow that it becomes blocked with powder or so wide that the resulting flame has a cold spot due to too large a central oxygen flow. The diameters of the outer nozzles limit the width of the flame. The feed material comprizes a specially prepared mixture of oxide components contained in a canister having a perforated bottom, for example a sieve, which can be vibrated with varying amplitude.

The feed powder falls into a conical hopper and thence via the central tube of the burner to the flame, where it impinges on the top of an alumina rod. Initially, by adjusting the gas flow rates, the flame is made sharp and of relatively low temperature, so that a sintered cone builds up on the alumina rod. The tip of this cone is melted by increasing the flow of coal-gas or hydrogen to the flame, and as more powder falls upon the cone the alumina rod is lowered at a uniform slow rate, approximately one centimetre per hour, so that the surface of the molten cap supported on the cone remains at a fixed position in the flame. In this way solidified material is gradually withdrawn from the bottom of the liquid cap, and if the original cone is sufficiently narrow so that only one nucleus forms the resulting solid is a single crystal. By further increasing the flame temperature the 'boule', as the crystal is called, may be widened to a diameter of a centimetre or so. It is possible also to begin growth of the boule from a seed crystal, the product of an earlier run, by cementing the seed on the alumina pedestal prior to the growth, and melting the top of the seed before switching on the powder flow. This dispenses with the careful 'necking' operation at the beginning of growth and also enables oriented crystals to be grown.

For the success of the flame fusion method it is essential that the feed material should have a loose structure and low bulk density so that it will react readily in the flame. Suitable oxide powders may be prepared by calcining in air, at temperatures up to 1000°C, compounds of the appropriate metals which lose a considerable proportion of their atomic constituents in the process. Such compounds are the double sulphates of the metals with ammonium which, in crystalline form, also contain a large amount of water. These have been used extensively since the invention of the process. Where a mixture

of oxides is required the double salts of the metals may be mixed in an end runner mill before calcination, or co-crystallized. More recently, W. H. BAUER *et al.*[10] have calcined the mixed oxalates of the metals, first prepared by co-precipitation from the metal chlorides. This technique requires care,

Figure 2. Photograph of the flame fusion apparatus

due to differences in solubility of the co-precipitated compounds, if the feed materials are to contain the correct proportions of the constituent oxides. G. ECONOMOS[14] has also prepared feed materials by the combustion in an atomizing burner of a solution of the appropriate metal nitrates in alcohol. In our laboratory the oxide powders, after preparation, are passed through a 300 mesh and kept dry in an oven at 80°C.

A photograph of the apparatus used in these laboratories is shown in *Figure 2*. To promote powder flow the hopper is maintained free from moisture by being kept at 100°C using a small heater and its inside surface is polished before use. An additional vibrator is applied to the neck of the hopper to prevent the build-up of powder. Even so, powders with a high proportion of free Fe_2O_3 tend to stick and flow unevenly. A subsidiary electric furnace is placed inside the main structure (*Figure 1*) for the adjustment of temperature gradients. The apparatus has facilities for continuous rotation of the crystal during growth. It is possible that rotation may produce a more uniform boule. However, it is important that the rotation, burner and lowering axes be collinear.

A typical boule with composition $MnFe_2O_4$, grown in the above apparatus is shown in *Figure 3*. We have found that the crystal growth rate should not exceed seven millimetres per hour. Much faster growth results in insufficient time for dissolved gas, from the flame, or from the dissociation of Fe_2O_3, to escape from the molten cap prior to crystallization. The gas pockets resulting from fast growth frequently result in the growth of additional crystallites, and the boule becomes polycrystalline.

Even when a subsidiary electric furnace is used the temperature gradient down the boule is so steep (1600°C at the top down to 1100°C at the bottom) that cracking tends to occur. This can be reduced by limiting the diameter of the boules.

A further cause of cracking is the superficial re-oxidation which occurs subsequent to crystallization as the surfaces of the boule cool. This is particularly marked in materials containing a large proportion of iron oxide. In such compositions the solid immediately after crystallization is in a partially reduced state, containing some ferrous iron. At a lower temperature the flame atmosphere becomes more oxidizing and the surface layer of the crystal takes up oxygen with the change of ferrous to ferric iron, and in some cases the ex-solution of an oxygen-rich phase such as α-Fe_2O_3. Such a surface layer will in general have a coefficient of thermal expansion different from the interior and differential stresses will be set up, with the possibility of cracking.

Figure 3. Single crystal boule of manganese ferrite grown by the Verneuil process

The surface layer may be detected by microscopic observation of polished surfaces and by resistivity measurements, using a two or four-point probe method. Thus the resistivity of the interior of a boule of nominal composition $MnFe_2O_4$ is 50 ohm cm approximately whilst its surface gives a value

of 50,000 ohm cm. Similarly, a boule grown from feed powder with composition Fe_2O_3 has internal resistivity 3·5 ohm cm and surface resistivity 1500 ohm cm. A polished cross-section of this boule is shown in *Figure 4*. The re-oxidized layer of α-Fe_2O_3 is apparent. Under higher magnification the Widmanstätten precipitates within the central core are also visible.

It has not been found possible in these laboratories to grow satisfactorily single crystals of the manganese oxides by the Verneuil method. The flame atmosphere is such that Mn_3O_4 is the stable phase at the melting point, and cubic Mn_3O_4 solidifies as the melt cools. At 1160°C, however, a phase transition from the cubic to the tetragonal form occurs and the solid becomes multiply twinned.

Attempts to grow the Ferroxdur composition $BaFe_{12}O_{19}$ have resulted in a foliated mass of partially oriented crystals. The growth habit of this hexagonal material is lamellar with the *c*-axis normal to the plate, so that the Verneuil method produces boules containing small crystals with their *c*-axes at right angles to the boule growth direction.

The application of the Verneuil method to incongruently melting compounds has been referred to by Bauer *et al.*[6] and by Rudness and Kebler[13]. In their synthesis of mullite $3Al_2O_3.2SiO_2$, which melts incongruently to give solid corundum in equilibrium with liquid of a higher silica content, Bauer and his colleagues arranged that the temperature gradient over the molten cap should be so steep that the liquid supercooled from the liquidus

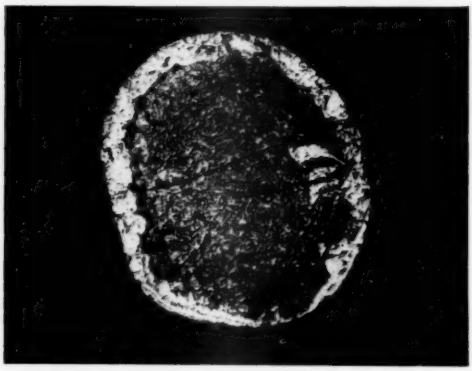


Figure 4. Polished cross-section of a boule grown from iron oxide powder

temperature and precipitated mullite as the solid phase rather than corundum. By contrast, Rudness and Kebler, in their growth of yttrium iron garnet, used very slow growth rates, crystallizing the solid under equilibrium conditions. They explained their technique on the basis of the phase diagram obtained by J. W. Neilsen and E. F. Dearborn[15], *Figure 5*. As the liquid of composition $Y_3Fe_5O_{12}$ cools the first phase to precipitate is $YFeO_3$, during which stage the molten cap becomes enriched in

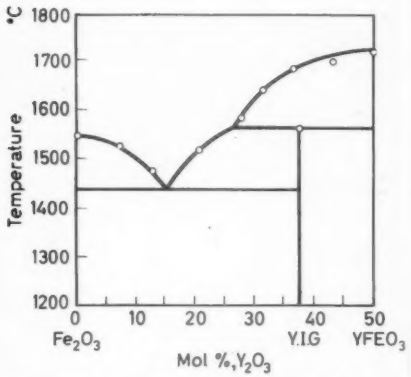


Figure 5. Proposed phase diagram for Fe_2O_3 and $YFeO_3$ (after Nielson and Dearborn)

Fe_2O_3 and the liquidus temperature falls until the liquid reaches the Fe_2O_3-rich peritectic composition. At this point yttrium iron garnet begins to separate at the bottom of the melt, whose composition is maintained by the addition of an equal quantity of feed powder to the top of the cap. At the end of the run, the molten cap is quenched, solidifying with the peritectic composition. Thus, if the top and bottom of the boule are discarded the remainder is an yttrium iron garnet single crystal. Rudness and Kebler did not say whether their crystals possessed the stoichiometric oxygen content.

The composition grown most satisfactorily in this laboratory has been NiO. The boule surface does not change in composition appreciably on cooling from the melting point; being cubic it is not subject to anisotropic stresses, and the result is a single crystal which is uniform and crack-free.

Bridgman–Stockbarger Method

The Bridgman–Stockbarger method is well known as a standard method for the growth of metal and other crystals. The material to be grown is contained in a cylindrical crucible whose length is several times its diameter and which is pointed at its tip. It is suspended in the hot zone of a vertical electrically heated tube furnace and the temperature raised until the material is completely melted. The temperature is then controlled whilst the crucible is lowered slowly to a cooler part of the furnace.

Crystallization begins at the tip of the crucible where the probability of formation of only one nucleus is high. This serves as a seed for the crystallization of the remainder of the melt.

J. SMILTENS[16] utilized the method for the growth of mechanically sound, stoichiometric single crystals of magnetite, Fe_3O_4. He employed a platinum crucible and a tubular glowbar furnace within which a small 'booster' winding of platinum supplied the high melting temperature and thermal gradient. His work was facilitated by the existence of very complete phase data for the iron–oxygen system[17,18], and the fact that magnetite, in contrast to the other ferrites, has an equilibrium oxygen pressure at its melting point considerably less than one atmosphere, 10^{-2} atmospheres approximately. He obtained the low oxygen pressure atmosphere around his crucible by making use of the known dissociation data for carbon monoxide–carbon dioxide mixtures. Not only did he arrange that the oxygen pressure at the melting point was correct, but he also adjusted the ratio of carbon monoxide to carbon dioxide in the furnace tube to maintain an equilibrium atmosphere as the temperature was lowered.

D. G. WICKHAM[19] attempted to grow single crystals of $CoFe_2O_4$ by the Bridgman method but his samples consisted of a number of large grains, were deficient in cobalt, and were non-stoichiometric with respect to oxygen. He concluded that it would be necessary to prevent the loss of cobalt oxide by evaporation and to use oxygen pressures much greater than one atmosphere if a stoichiometric ferrite were to be obtained. This finding was confirmed[20] in an investigation of the ferrite region of the phase diagram Fe–Co–O.

L. R. BICKFORD, J. M. BROWNLOW and R. F. PENOYER[21,22] succeeded in growing magnetite, and magnetite containing small amounts of cobalt up to $Co_{0.1}Fe_{2.9}O_4$, using a pure carbon dioxide atmosphere for magnetite and adding increasing small amounts of oxygen as the cobalt content was raised. Unlike Smiltens, they did not vary the oxygen content of the atmosphere on cooling the crystals, but partially quenched them after solidification by rapid lowering out of the hot zone of the furnace. The crystals, contained in platinum–10% rhodium crucibles $2\frac{1}{4}$ in. long and $\frac{5}{16}$ in. in diameter, showed no apparent oxidation or mechanical weakness as a result of this treatment. There was variation of cobalt content along the crystal, increasing from the bottom upward.

R. F. PENOYER and M. W. SHAFER[23] grew crystals of composition $Mn_xFe_{3-x}O_4$ with specific compositions having x varying from 0·10 to 1·90, and a metal to oxygen ratio of three to four. The crystals were grown at a rate of 1°C per hour and quenched to room temperature after crystallization at a rate of 100°C per minute approximately. The underlying data for this growth programme were obtained from Shafer's phase equilibrium investigation[24] at liquidus temperatures in the Mn–Fe–O system.

In our laboratories the method has been used by R. CRUMPTON for the growth of Fe_3O_4, $Fe^{2+}Fe^{3+}_{2-x}Al_xO_4$ and $Fe^{2+}Fe^{3+}_{2-x}Ga_xO_4$ in connection with measurements of magnetocrystalline anisotropy by R. F. PEARSON[25]. The value of x varied from zero to 0·2 for the composition containing aluminium and from zero to 0·8 for that containing gallium. A schematic diagram of the furnace used is shown in *Figure 6*. This possesses a molybdenum winding,

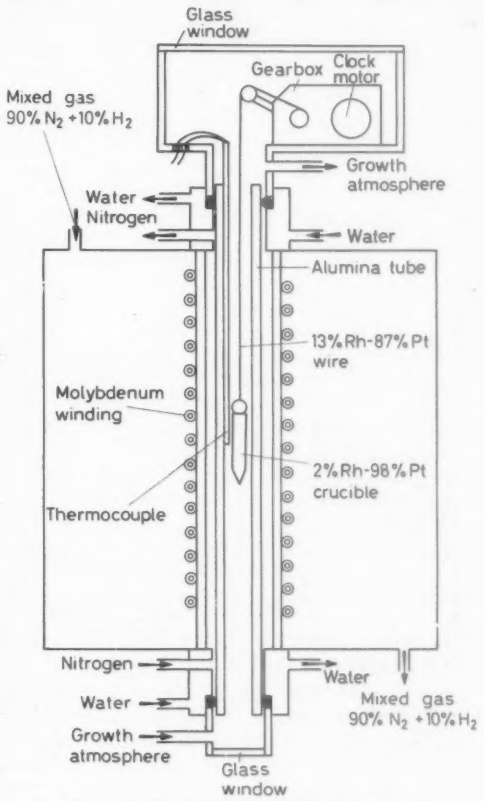


Figure 6. Typical Bridgman furnace layout

which whilst requiring to be protected by a reducing atmosphere, in this case mixed gas, enables high temperatures to be reached without the risk of element failure involved in the use of platinum and

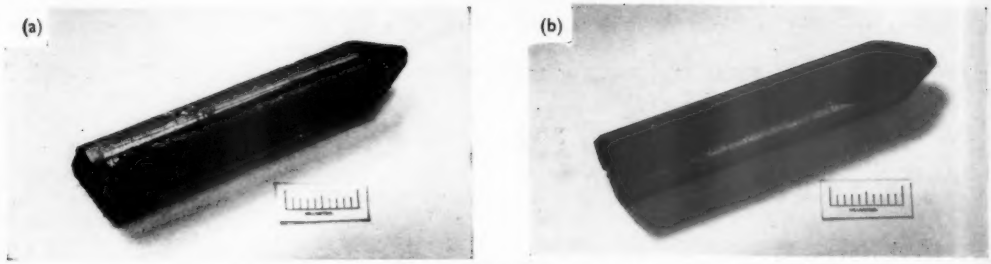


Figure 7. Magnetite crystal grown by the Bridgman method etched in hydrochloric acid (a) showing re-oxidized layer (b) showing crystal boundary

low rhodium-content platinum alloys. A second alumina tube contains the growth atmosphere and the annular space between the two tubes is flushed continuously with nitrogen to prevent leakage of hydrogen from the furnace casing into the inner tube. The end seals of the inner tube are cooled with brass water jackets. The crucibles are 10 cm long and 1 cm in diameter.

In the preparation of the above compositions, the mixed constituent oxides were first pressed into rods under a pressure of 15 ton/sq.in. using distilled water as a binder and sintered for five hours at 1500°C in carbon dioxide. Part of such a rod was ground and the powder produced packed into the cone of the crucible, the remainder of the rod being then inserted into the crucible. It was found necessary to pack the crucible tip carefully, since on occasion samples after growth consisted of a number of large grains which could be associated with the trapping at the tip of a bubble of gas. In addition, crucibles were not filled more than two-thirds full to avoid the possibility of running over of the melt during the time that gas was evolving as the material came to equilibrium with the furnace atmosphere. During growth, the crucibles were lowered at a uniform rate of 2 cm per hour through a temperature gradient of 10°C per centimetre giving a cooling rate of 20°C per hour, somewhat faster than that used by Smiltens (10°C per hour). The crystals were extracted after growth by spirally stripping the platinum–rhodium from them.

Carbon dioxide was used as a growth atmosphere for all the above compositions. Quenching of the type described by Bickford *et al.*[21] was not adopted since it was found that cracking tended to occur, probably because our crystals were larger than those of these authors. However, it was found that only the surfaces of the crystals in contact with the atmosphere suffered slight reoxidation when a slower cool was carried out. The remaining surfaces were protected from re-oxidation by the crucible. *Figure 7* shows such a crystal of magnetite after etching for two minutes in cold concentrated hydrochloric acid. In *Figure 7a* the re-oxidized layer at the top of the crystal is clearly seen, whilst in *Figure 7b* the boundary between the main crystal and a subsidiary which formed at the beginning of the growth is delineated.

Chemical analyses for iron, gallium, aluminium and oxygen deficiency or excess on three typical crystals give the following formulae, referred to four oxygen atoms:

Top of crystal, excluding re-oxidized surface	*Bottom of crystal*
$Fe^{2+}_{0\cdot97_0}Fe^{3+}_{2\cdot02_0}O_4$	$Fe^{2+}_{0\cdot94_3}Fe^{3+}_{2\cdot03_8}O_4$
$Fe^{2+}_{0\cdot97_2}Fe^{3+}_{1\cdot63_2}Ga_{0\cdot38_7}O_4$	$Fe^{2+}_{0\cdot92_9}Fe^{3+}_{1\cdot60_6}Ga_{0\cdot44_2}O_4$
$Fe^{2+}_{0\cdot98_4}Fe^{3+}_{1\cdot80_2}Al_{0\cdot20_9}O_4$	$Fe^{2+}_{0\cdot93_1}Fe^{3+}_{1\cdot84_7}Al_{0\cdot19_9}O_4$

All the compositions show a slight oxygen excess which may be due to the presence of a trace of oxygen in the carbon dioxide atmosphere. It is not due to re-oxidation at a lower temperature since if this were so the top of the crystal would be expected to contain less Fe^{2+} than the bottom.

With the relatively slow cool used for these compositions it is found that crystals containing greater than 0·2 Al or 0·8 Ga per formula unit exhibit epitaxial precipitates, presumably of an oxygen-rich second phase. Quenching from below the melting point reduces this second phase, but crystals of such a size crack upon this treatment.

Flux Methods

As already stated, the growth of crystals containing ferric oxide as a component is made difficult because the melting points of such compositions are higher than the dissociation temperature of ferric oxide in oxygen at atmospheric pressure. This difficulty can be avoided if a suitable flux can be found, such that the system comprising the crystal composition and the flux possesses a eutectic with temperature lower than the dissociation point of ferric oxide. The growth rate for crystallization from a flux must

be much slower than for crystallization from a melt of the crystal composition if equilibrium conditions are to be maintained and the crystal is to be uniform. This is because the act of crystallization tends to change the composition of the liquid in the region of the crystal and time must be allowed for diffusion processes in the liquid to maintain uniformity. Similarly, it is not generally possible to use the Bridgman–Stockbarger method. It is usual, having completely melted the solution of flux and compound required in a crucible of conventional shape, to cool the furnace at a very slow uniform rate by careful control of the power input to the furnace. Crystals nucleate at random and can be leached out of the matrix at the end of the run with an appropriate solvent. Usually crystals are numerous and small.

In the growth of ferrite single crystals the first flux to be used was borax. Nickel ferrite[26], manganese–zinc ferrite[27,28], cobalt–zinc ferrite[29] and copper ferrite[30] were all grown from this flux. However, the largest crystals obtained were no more than two millimetres in linear dimensions, and were in the majority of cases not mechanically sound. J. K. GALT *et al.*[27] used melts containing approximately 16 g of borax to 10 g of the ferrite composition which they first heated for some hours at 1330°C in a platinum crucible, before cooling at rates as slow as 2°C per hour or as fast as 100°C per hour. T. OKAMURA *et al.* used a similar method[28], obtaining comparable results with the ferrite compositions which they attempted.

It is sometimes possible to find a composition suitable for crystal growth from solution by the addition of an excess of a component already present in the system. Thus, A. L. Stuijts (see C. KOOY[31]) prepared single crystals of $BaFe_{12}O_{19}$ of good quality. W. V. SMITH *et al.*[32] prepared single crystals of gadolinium iron garnet about one millimetre in linear dimensions by heating a mixture of 80 mol.% Fe_2O_3 and 20 mol.% Gd_2O_3 to 1520°C in oxygen and then slowly cooling. The Gd_2O_3–Fe_2O_3 system is somewhat similar to the Y_2O_3–Fe_2O_3 system[15] and the growth process of the gadolinium iron garnet may be understood by reference to *Figure 5*. Whilst a solid garnet melts incongruently to produce liquid and a solid *ortho*ferrite, a liquid with excess iron oxide such that the rare earth oxide content lies between 15 and 27 molar per cent will on cooling precipitate solid garnet alone until the eutectic temperature is reached. Further cooling solidifies the liquid of eutectic composition as a two phase mixture of Fe_2O_3 and garnet. It will be seen that the minimum temperature at which solidification occurs is 1440°C approximately, still relatively high, and it is likely that garnet crystals prepared in this way, even using an oxygen atmosphere, are somewhat oxygen-deficient.

The growth of garnet and other oxide crystals at lower temperatures has been made possible by workers at the Bell Telephone Laboratories (U.S.A.), through their investigations of lead oxide, PbO, as a flux. Spinel ferrites form a system with lead oxide having a eutectic at 800°C and approximately five per cent by weight ferrite composition[33]. Thus, a composition containing 30 per cent by weight of nickel ferrite and 70 per cent by weight of lead oxide is completely liquid at 1300°C. This composition is a convenient one from which to grow nickel ferrite crystals by slow cooling, since at 1300°C Fe_2O_3 does not dissociate in air, nor is the platinum crucible, in which the melt is contained, attacked to any extent by the molten oxides. Cooling of such a melt at 10°C per hour produces many small crystals which, whilst likely to contain inclusions, may be used as seeds on which more perfect growth can be carried out in subsequent runs. J. P. REMEIKA[34], having grown a number of spinels from lead oxide melts, applied the method to the growth of compounds of the type $A^{3+}B^{3+}O_3$ with Y, La, Pr, Nd, Eu, Sm and Gd in the A position and Al, Sc, Cr, Fe, Co and Ga in the B position. Subsequently, Nielson and Dearborn[15] examined the ternary phase diagram of the system PbO–Fe_2O_3–M_2O_3 where M is a rare earth ion or yttrium. They established a suitable melt composition for garnet growth as PbO 52·5, Fe_2O_3 44, M_2O_3 3·5 mol per cent where M is Y, Gd, Er or Sm. This composition is molten below 1350°C, and when cooled at 1°C per hour crystals weighing 2 g are quite common.

In addition to the growth of garnet crystals, crystals of magnetoplumbite $PbFe_{12}O_{19}$ separate from melts of the composition mentioned above. Indeed, the latter compound is the primary phase for such melts, and forms large intergrowths over the surface of the solidified melt. This is illustrated in *Figure 8a* where a mass of magnetoplumbite crystals from a garnet growth run by J. L. PAGE of these laboratories is shown. Garnet crystals are found attached to the underside of the magnetoplumbite mass, to the walls of the platinum crucible, or embedded in the lead oxide matrix. *Figure 8b* shows a crystal of yttrium iron garnet grown by Page, who has also grown gadolinium and ytterbium garnets. The possession of good external faces by a garnet crystal does not guarantee freedom from inclusions. A crystal similar to that of *8b*, when sectioned and polished, appeared as in *Figure 8c* when viewed under a metallurgical microscope. 'Skeletal' inclusions are present consisting of two phases, one darker and one lighter than the garnet.

(a)

(b)

(c)

Figure 8. Growth of garnet crystals from lead oxide solution—(a) Magnetoplumbite intergrowths formed on the surface of the melt; (b) Yttrium iron garnet crystal; (c) Polished cross-section of a similar crystal, showing inclusions

Grain Growth

In the metallurgical field the strain–anneal method of crystal growth, in which the pronounced growth of certain grains in a polycrystalline rod is induced by straining the rod and subsequently annealing at a higher temperature, is well known. In the field of refractory oxides this technique is not available, since it is not possible to induce preferred orientation by applied stress in such hard materials. However, pronounced grain growth in polycrystalline ferrites heated for some time at high temperatures has been observed. The occurrence of grains up to 5 mm in size has been described[35] in some polycrystalline yttrium iron garnet samples fired above 1450°C. This phenomenon has been noted in our laboratories in yttrium iron garnet and also in copper manganese ferrite samples sintered at 1400°C. Such marked grain growth is obviously connected with the mechanism of sintering, which is at present incompletely understood. It is possible that progress in this subject will make grain growth a practicable and attractive method of single crystal production.

Other Methods

Much research has been carried out by J. KOENIG[36] on the hydrothermal method of crystal growth applied to ferrites. Using aqueous solutions of ammonium chloride principally and magnetite in powder form or as pressed slugs, contained in a steel autoclave, he employed pressures of the order of 20,000 lb/sq.in. and temperatures in the range 400° to 500°C. In many cases spontaneous crystallization occurred resulting in crystals up to a millimetre in linear dimensions. He succeeded in producing growth on a magnetite seed at the rate of 0·05 mm per day. His experiments with spinel ferrites other than magnetite were less productive.

H. P. J. WIJN[37] used induction heating to prepare a number of ferrites. His starting composition contained the iron oxide as magnetite, which because of its relatively low resistivity, 10^{-2} ohm cm at room temperature, enabled eddy currents sufficient to heat the charge to be produced. The resistivity of the oxide composition as a whole was low enough at elevated temperatures for the melting point to be attained. The melt was slowly cooled by the gradual reduction of the high frequency current and crystals up to 5 mm in linear dimensions were grown. Wijn was able to use crucibles of high purity alumina which, being at a somewhat lower temperature than the melt, did not react appreciably with it. The crystals grown were magnetite, manganese ferrite and some of the barium compounds with hexagonal structure of which $BaFe_{12}O_{19}$ and $BaFe_{18}O_{27}$ are examples. Wijn pointed out that

the method could be applied to the growth of crystals under high pressures by enclosing the apparatus in a bomb.

More recently, R. E. CECH and E. I. ALESSANDRINI[38] have prepared single crystals of FeO, NiO and CoO as thin films on the cleaved surfaces of magnesium oxide single crystals. The oxides form when water vapour is passed over the halides of the metals in the presence of the magnesium oxide crystals at temperatures between 570° and 720°C as a result of the reaction

$$MX_2 + H_2O \rightarrow MO + 2HX,$$

where M is Fe, Co or Ni and X is Cl or Br. The metal halides vaporize at these temperatures and the magnesium oxide which, like FeO, CoO and NiO, has the rock salt structure, and a comparable lattice constant, acts as a catalyst, the oxides being deposited only on the magnesium oxide and not on all surfaces in the reaction chamber. Oxide crystal films up to 500μ thick have been produced in this way.

Conclusion

The foregoing description indicates the degree of success which has been achieved in the growth of oxide single crystals containing transition metal ions. Of the three main methods described, Verneuil, Bridgman–Stockbarger, and lead oxide solution, it cannot be said that any one is universally more advantageous than the others. The Bridgman–Stockbarger method yields the most mechanically sound crystals, but like the Verneuil method does not permit the preparation of stoichiometric crystals with high equilibrium oxygen pressures. The Verneuil method is probably the most economical and dispenses with the need for a crucible. However, it requires the attention of a skilled operator and the crystals produced are likely to be strained. The lead oxide method is most promising in view of the low growth temperature, but a crystal growth experiment is long, approximately two weeks, and very critical temperature control is required throughout that time. Furthermore, the crystals produced until now are small and require careful selection if inclusions are to be avoided.

J. L. Page and R. Crumpton grew the crystals from which many of the illustrations have been produced and J. Dale prepared the photomicrographs. Figure 5 is reproduced by courtesy of the Editor of the Journal of the Physics & Chemistry of Solids. *I am indebted to the management of the Mullard Research Laboratories for permission to publish this paper.*

References

1 WENT, J. *et al. Philips Tech. Rev.* (1952) **13**, 194
2 JONKER, G. H., WIJN, H. P. J. and BRAUN, P. B. *ibid.* (1956) **18**, 145
3 VERNEUIL, M. A. *C. R. Acad. Sci., Paris* (1902) **135**, 791
4 MERKER, L. *FIAT Final Report No.*1001 (March, 1947)
5 MOORE, C. H. *Mini. Engng* (1949) **1**, No. 6, Sect. 3, 194
6 BAUER, W. H., GORDON, I. and MOORE, C. H. *J. Amer. ceram. Soc.* (1950) **33**, 140
7 — and KEAT, P. P. *Ceram. Age* (1950) **56**, 18
8 — and GORDON, I. *J. Amer. ceram. Soc.* (1951) **34**, 250
9 — *Ceram. Age* (1954) **63**, 41
10 — MERZ, K. M. and BABA, P. D. *Tech. Rep. No.* 1 (June, 1958) *ONR Contract No. Nonr-404 (oz) NR* 032–347
11 SCOTT, E. J. *J. chem. Phys.* (1955) **23**, 2459
12 POPOVA, A. A. *Sov. Phys. Doklady* (1958) **3**, 711
13 RUDNESS, R. G. and KEBLER, R. W. *Amer. ceram. Soc. Bull.* (April 1959), p. 169
14 ECONOMOS, G. *Progress Report No.* XXIII (June 1958), p. 54, Laboratory for Insulation Research, Massachusetts Institute of Technology
15 NIELSON, J. W. and DEARBORN, E. F. *J. Phys. Chem. Solids* (1958) **5**, 202
16 SMILTENS, J. *J. chem. Phys.* (1952) **20**, 990
17 DARKEN, L. S. and GURRY, R. W. *J. Amer. chem. Soc.* (1945) **67**, 1398
18 — — *ibid.* (1946) **68**, 798
19 WICKHAM, D. G. *Tech. Report* 89 (1954) Laboratory for Insulation Research, Massachusetts Institute of Technology
20 SMILTENS, J. *Tech. Report* 115 (1957) Laboratory for Insulation Research, Massachusetts Institute of Technology
21 BICKFORD, L. R., BROWNLOW, J. M. and PENOYER, R. F. *Proc. Inst. elect. Engrs* (1957) **104B**, Suppl. 238
22 — — — *J. appl. Phys.* (1958) **29**, 441
23 PENOYER, R. F. and SHAFER, M. W. *J. appl. Phys. Suppl.* (April 1959) **30**, No. 4, 315*S*
24 SHAFER, M. W. *I.B.M. J. Res. Dev.* (1958) **2**, 193
25 PEARSON, R. F. *Colloque International de Magnetisme, Grenoble* (July 1958), p. 345
26 GALT, J. K., MATTHIAS, B. T. and REMEIKA, J. P. *Phys. Rev.* (1950) **79**, 391
27 — *et al. ibid.* (1951) **81**, 470
28 OKAMURA, T., KOJIMA, Y. and TORIZUKA, Y. *ibid.* (1952) **85**, 693
29 — — *ibid.* p. 690
30 — — *ibid.* (1952) **86**, 1040
31 KOOY, C. *Philips Tech. Rev.* (1958) **19**, 286
32 SMITH, W. V., OVERMEYER, J. and CALHOUN, B. A. *I.B.M. J. Res. Dev.* (April 1959) **3**, No. 2, 153
33 Western Electric Co. Inc. Brit. Pat. Specn. 783,670
34 REMEIKA, J. P. *J. Amer. chem. Soc.* (1956) **78**, 4259
35 VAN UITERT, L. G., SWANEKAMP, F. W. and HASZKO, S. E. *J. appl. Phys.* (1959) **30**, 363
36 KOENIG, J. *U.S. Govt. Research Reports*, PB 112795, PB 114024, PB 114669, PB 115923, PB 116516, PB 117348, PB 117723, PB 111934 (Contract AF 19 (604)—857), PB 122360, PB 122191, PB 123450, PB 124140 (Contract AF 19 (604)—1419)
37 WIJN, H. P. J. *Nature, Lond.* (1952) **170**, 707
38 CECH, R. E. and ALESSANDRINI, E. I. *Trans. Amer. Soc. Metals* (1959) **51**, 150

THE MEASUREMENT OF OPTICAL RADIATION

E. J. GILLHAM

Light Division, National Physical Laboratory

Methods used in a standardizing laboratory for making absolute measurements of radiation in the ultra-violet, visible and infra-red, are described, emphasizing mainly radiation detectors of the thermal type, including absolute radiometers, and how they are used either for direct measurement, or for calibration of other detectors.

THE measurement of radiation in the ultra-violet, visible and infra-red regions of the spectrum is not only of intrinsic scientific importance, but also plays a part in many laboratory and industrial techniques. In recent years there has been considerable development in the practical application of radiation measurement, and in methods of detection and measurement suitable for routine work. However, it is not the purpose of this article to discuss these, but rather the basic methods of measurement and calibration which form an essential background to these practical applications.

In principle, the measurement of radiant energy requires the conversion of the energy into some other measurable form with predictable efficiency. The various photochemical or photoelectric processes of conversion which are utilized for the detection of radiation are only partially efficient and are, moreover, wavelength-dependent or selective. The only process of predictable efficiency is that of degradation into heat. Hence the measurement of radiation depends ultimately on detectors of the thermal type, in which the radiation is absorbed at the blackened surface of a thermally isolated receiver, the rise in temperature of which is indicated by some temperature-sensitive element.

Thus this article will be concerned mainly with detectors of this kind, and the way in which they are used, at a standardizing laboratory such as the National Physical Laboratory, either for direct measurement, or for the calibration of other more sensitive or convenient detectors. This process of establishing a scale of radiant energy may be regarded as taking place in two stages, the first of which is concerned with the problem of selectivity. The sensitivity of a thermal detector is proportional to the absorption factor of the black with which it is coated, and this may vary by several per cent over the spectrum. The first stage therefore consists in determining the selectivity of a suitable detector by either measuring or controlling its absorption factor in some way. When this has been done, the means exist for giving equal weight, on an energy basis, to radiation of any wavelength, and hence for measuring radiation of any spectral quality on an arbitrary but consistent scale of energy. The second stage consists in relating the units of this scale to the accepted units of energy, in other words, of making absolute measurements of radiant energy.

The quantities which a radiation measurement is designed to yield must here be defined. Excluding questions of polarization or spectral distribution, the radiation field at a particular point is completely specified by the radiation brightness, or steradiance, as a function of direction. In many cases, however, it is sufficient to measure the intensity, *i.e.*, the radiant power crossing unit area from all directions. This definition is arbitrary to the extent that it involves the orientation of the unit area. However, if this is defined by the unit vector $\mathbf{n}$ normal to the surface, then the measured intensity I_n is given by

$$I_n = \mathbf{I} \cdot \mathbf{n} \text{ where } \mathbf{I} = \int \mathbf{B}\, d\omega,$$

$\mathbf{B}$ being the brightness regarded as a vector quantity and ω solid angle. Thus the intensity vector $\mathbf{I}$ is a uniquely defined property of the radiation, and is the appropriate quantity to measure when the direction $\mathbf{n}$ is not otherwise determined. It should be noted, however, that intensity is not always the quantity of greatest practical significance, and that in some applications, *e.g.* in measuring the solar irradiation of plants, it may be more appropriate to measure the scalar integral $\int \mathbf{B}\, d\omega$ which is equivalent to radiant energy density.

In practice, however, it is only necessary to distinguish between two types of measurement. The first is concerned with the intensity at a particular point in an extended field of radiation, generally of uniform intensity, and the second with the total radiant power in a beam of limited cross-section. The measurement of brightness is sometimes required, but this differs from the measurement of intensity only in that the angle of view of the radiometer is suitably limited by diaphragms.

Measurement of Intensity of Total Radiation

The technique of radiation measurement may be further classified according to whether the radiation to be measured is monochromatic or not. Since monochromatic radiation is usually got by isolating a narrow region of a continuous spectrum with a monochromator, the radiant power available is small (a microwatt or less), and it may be difficult to attain the necessary precision owing to fluctuations in the receiver or associated electrical circuits. This difficulty can, however, be overcome in many parts of the spectrum by using a highly-sensitive, but selective, detector, calibrated at the required wavelengths. On the other hand, in measuring 'total' radiation, *i.e.* radiation with a continuous spectrum, the radiometer must be non-selective, and is therefore almost invariably an instrument of the thermal type. The inferior sensitivity of such instruments is not necessarily a drawback, since the radiant power available is usually quite high, a milliwatt or more.

The type of measurement most frequently encountered in practice is the measurement of the intensity, at a particular point, of the total radiation emitted by some kind of thermal radiator, such as a furnace or a tungsten-filament lamp. The instrument most commonly used for this purpose is the radiation thermopile of which a typical example is the Moll thermopile. In this, about twenty thermocouples, each in the form of a thin, flat strip, with the thermojunction in the centre, are soldered to insulated pins, so that they are stretched parallel to each other, with small gaps in between, and cover an area of about 8 mm × 6 mm. The insulated pins are in good thermal contact with the body of the thermopile, and act as cold junctions. The surface of the strips is blackened, usually with an evaporated metal black, and the area exposed to radiation is limited by a circular diaphragm six millimetres in diameter which shields the cold junctions from the incident radiation. The thermocouples are all connected in series, and have a total resistance of 20 to 30 ohms.

The Moll thermopile gives a steady response in little more than a second after exposure to radiation, and there is no trace of the slow downward drift in reading sometimes exhibited by thermopiles owing to heating up of the cold junctions. The response is not exactly proportional to radiation intensity, the sensitivity falling linearly by about 3 per cent per W/cm², but since the thermopiles are rarely employed at intensities greater than 0·1 W/cm², this effect can usually be neglected. In many practical applications, the thermopile output is indicated on a moving-coil meter, but for calibration and other precise work, it is usual to measure the open-circuit voltage with a galvanometer and potentiometer. In terms of open-circuit voltage, the sensitivity is of the order of 0·07 V/W/cm². The angle of view of the thermopile is usually limited by a diaphragm to about 25° from the normal. However, at this angle of incidence, the thermopile response may depart from the cosine law by 5 per cent or more, and a correction should therefore be applied if the radiation source subtends an angle of more than a few degrees.

The Moll thermopile is sensitive, not only to draughts, which can be eliminated by boxing-in the apparatus, but also to atmospheric pressure fluctuations caused, for example, by boisterous weather or slamming doors. This effect can be avoided by mounting the thermopile in a hermetically sealed enclosure provided with a window, but this restricts the spectral range of the instrument, and for most purposes it is desirable to operate the thermopile in air, without windows. It is then often advantageous to connect a similar, compensating thermopile in opposition to the one exposed to radiation. The output fluctuations in favourable conditions are then of the order of 0·1 μV corresponding to a radiation intensity of the order of 1 μW/cm².

An air operated thermopile is sensitive to room temperature radiation and will therefore be affected by any temperature difference between the thermopile and its surroundings. The temperature difference may be only a fraction of a degree, very small compared with the temperature of the source which provides the radiation that is to be measured. On the other hand, the surroundings visible to the thermopile usually subtend a much greater solid

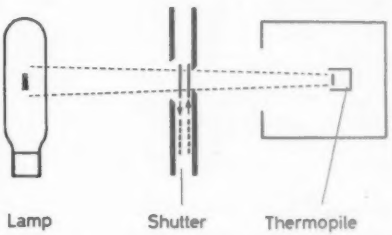


Figure 1. Shutter arrangement for total radiation measurement

angle than the source, so that the zero shift of the response caused by room temperature radiation may be an appreciable fraction of the desired response. This effect can be overcome by restricting the angle of view of the thermopile so that it is just

sufficient to embrace the whole of the source, using a diaphragm in good thermal contact with the body of the thermopile. It is usually more convenient, however, to employ a shutter which can cut off the radiation from the source, and to take as the true response of the thermopile, the difference between the readings obtained in the open and closed position of the shutter. An additional advantage of this method is that it effectively eliminates parasitic e.m.f.'s in the electrical measuring circuits.

Figure 1 shows a suitable shutter arrangement. The operation of the shutter should affect as little as possible the room radiation received by the thermopile, and it is therefore placed behind a screen with an aperture in it just big enough for the source to be visible from the thermopile. This screen is shielded from the source by one or more additional screens, provided with somewhat larger apertures, in order to prevent it from being heated by the radiation and so producing an inconveniently large zero reading. It is also desirable to prevent the shutter from heating up, even though it may only subtend a small angle at the receiver, and it is therefore constructed of two or three separate leaves thermally insulated from each other.

In some practical applications, it may be required to measure the radiant output of a low temperature radiator, operating, say, at a temperature T_s of 100°C or less. Since the measured radiation

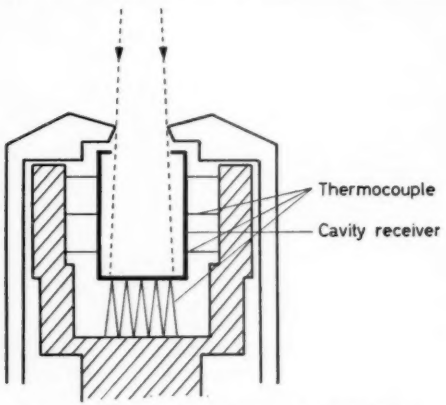


Figure 2. Black-body radiation for selectivity measurements

is proportional to $T_s^4 - T_0^4$ where T_0 is the temperature of the thermopile, it is then necessary to define T_0 with an accuracy of a degree or so, in accordance with the practical significance of the measurement. The shutter must then be maintained at this assigned temperature. If, as is most likely, the radiator is large, it may be more convenient to dispense with a shutter, and either control the temperature of the thermopile or alternatively obtain a zero reading by directing the thermopile towards a blackened surface maintained at the reference temperature T_0.

If the thermopile is hermetically sealed its output is much more stable, and it becomes possible to measure intensities of 1 $\mu W/cm^2$ or even less. If the window is of glass, quartz, fluorite or other material with a fairly short wavelength cut-off in the infra-red, room radiation ceases to be troublesome, although if low intensities are to be measured, it is still advisable to use a simple shutter to overcome the effect of parasitic e.m.f.'s. Although the precision of measurement is improved by sealing, care must be taken to ensure that, in using the instrument, the accuracy is not impaired by selectivity effects due to the window, or by directional effects due to the inter-reflection of radiation between the window and the thermopile.

Selectivity of Moll Thermopiles

Moll thermopiles, as normally supplied, are coated with a metallic black and are usually quite non-selective. For accurate work, however, the selectivity must be examined, and the way in which this is done forms a suitable introduction to the general problem of selectivity already referred to. In principle, selectivity may be measured by determining the reflectivity (and hence absorptivity) of a suitable thermopile or other thermal detector. The measurement of diffusely reflected radiation is rather difficult, however, particularly in the infra-red, and at the N.P.L. an alternative approach has been adopted, that of constructing a special thermopile with a 'black-body' or cavity receiver, which is almost completely absorbing, and therefore non-selective, over a wide spectral range.

Figure 2 shows the essential features of one such instrument. The receiver consists of a cylindrical can 1·5 cm long and 1 cm diameter, with an entrance hole at one end 8 mm in diameter. An external diaphragm 6 mm in diameter limits the incident radiation so that it falls only on the closed end of the can. The matt black paint with which the interior is coated has a reflection factor which does not exceed about five per cent over the spectral range concerned, and since only about one tenth of the radiation diffusely reflected from the end of the can escapes through the entrance hole, the effective absorption factor of the receiver is always within half a per cent of unity.

It is essential for the correct operation of the instrument that the sensitivity should be uniform over the whole of the interior surface, in other words,

that a given radiant flux should produce the same response in the temperature-sensitive element, irrespective of where in the receiver it is absorbed. To a large extent this is assured by thermal conduction through the walls of the receiver, which are

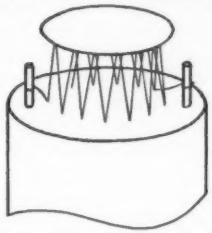


Figure 3. Circular target thermopile for measurement of radiant power

of $\frac{1}{2}$ mm thick aluminium, but as a further precaution the thermocouples which form the temperature-sensitive element are distributed over the surface of the receiver so as to average its rise in temperature. The thermocouples, about 40 in number, are made by soldering 40 SWG nichrome and eureka wire, and are cemented to the can with *Durofix*. The outer surface of the can is anodized to provide electric insulation.

Since the total surface area of the receiver is considerably greater than for an ordinary thermopile of similar receiving area, the heat losses are greater, and the sensitivity is only about half that of the Moll thermopile. Moreover, the thermal time constant is about ten seconds, and an exposure time of one minute is required for a steady reading. These disadvantages, however, are not serious in calibration work, although they would be troublesome in routine measurements.

It is the practice at the N.P.L. to examine Moll thermopiles for selectivity by calibrating them against a black-body radiometer, using four different types of radiation representative of the range of spectral qualities normally encountered in total radiation measurements. The radiation sources are: (*i*) a black-painted radiator operated at a temperature of 200°C; (*ii*) a nichrome wire or strip radiator operated at 1000°C; (*iii*) a tungsten filament lamp operated at a colour temperature of 2850°K; and (*iv*) the same lamp provided with a water filter and heat-absorbing glass to absorb the infra-red.

A good thermopile will not vary in sensitivity by more than 2 per cent or 3 per cent over this range of spectral qualities, so that when it has been calibrated for selectivity, it may be used to measure almost any type of total radiation with a selectivity error not exceeding about $\frac{1}{2}$ per cent.

With careful handling, a Moll thermopile will remain constant in sensitivity to within 1 or 2 parts per thousand for a year or more. It thus forms a convenient working-standard for maintaining the scale of radiation intensity set up by the absolute measurements later to be described, and also for extending the range of these measurements with regard to intensity and spectral quality. At the N.P.L. a set of six such thermopiles are used for this.

Measurement of Radiant Power

If it is required to measure the total power in a beam of limited cross-section, rather than the intensity in an extended field, the thermopile receiver must collect all the radiation. It cannot therefore be uniformly illuminated, except in the special case when the beam just fills the receiver. It is therefore essential that the sensitivity of the thermopile should be uniform over the receiving area. This is not the case with the Moll thermopile, for not only is the receiving surface not continuous, but the sensitivity falls off considerably towards the cold junctions at the side.

The conditions for uniform sensitivity are that the heat losses from the receiver should be uniform over its surface, and that the temperature-sensitive element should average the temperature rise of the receiver; since these conditions cannot be realized exactly, some reliance must be placed on the lateral conduction of heat across the receiver, which must therefore consist of a suitably thick gauge of high conductivity material. *Figure 3* shows one suitable form of construction. The receiver, which consists of a circular aluminium disc 1 cm in diameter and 0·1 mm thick, is cemented to an array of 24 thermocouples distributed uniformly over a circular area of about 9 mm diameter. The thermocouples are prepared by soldering 6 mm lengths of 40 SWG nichrome and eureka wire, and the cold junctions are cemented to an aluminium base plate. The thermocouples are connected in series, and the receiver and base plate are anodized to provide the necessary electrical insulation.

The thermal time constant of this instrument is quite long, owing to the high thermal capacity of the receiver, and about half a minute is required for a steady reading. The resistance is about 20 ohm. The sensitivity is uniform to within one per cent over the whole receiving area, and has a value of about 0·3 volts per watt. The instrument is calibrated in terms of radiant power by limiting the exposed surface of the receiver with a diaphragm of known area, and then calibrating in a uniform field of radiation against the Moll thermopiles which form the working standards of intensity.

There is available commercially a thermopile with a circular target, 2 mm in diameter, which is suitable for measuring total flux when the radiation can be focussed on to an area of this size. The thermopile is of the Schwarz type, in which the target is mounted on two pins tipped with semiconducting material, and itself forms part of the thermojunction. The target is normally made from very thin foil in order to obtain a rapid response, and the sensitivity is not very uniform. The instrument can, however, be supplied with a thicker target, which gives a sensitivity uniform to within two per cent, without increasing unduly the response time. In order to utilize the Schwarz type of construction most effectively, it is preferable to operate the thermopile in vacuum, using a window of fluorite or other material suitable for the intended spectral range. The sensitivity for a 150 ohm thermopile is then about 18 volts per watt, and since the response is unaffected by atmospheric disturbance, radiant powers down to about 0·01 μW can be measured.

It is not easy to calibrate the vacuum thermopile by direct comparison with intensity standards, since the window makes it difficult to mount an external diaphragm sufficiently close to the target, and also leads to error due to the inter-reflection of radiation. It is therefore preferable to calibrate the thermopile against an air-operated instrument, such as that described above. Since the response is proportional to radiant power over a wide range, this can safely be done at a radiant power level suitable for the less sensitive instrument.

Measurement of Monochromatic Radiation

We turn now to the thermopiles used for the measurement of monochromatic radiation. These play an important part in routine measurement, as well as in standardizing work, for although it is sometimes necessary, with monochromatic radiation, to use a highly-sensitive photoelectric detector, a thermopile is more stable in sensitivity, and is to be preferred when sufficient energy is available. Moreover, the thermopile still remains one of the most convenient and sensitive instruments for radiation of wavelength longer than about six microns.

A thermopile widely used for monochromatic measurements is the linear Schwarz thermopile, which is a vacuum instrument with a strip-shaped receiving element. Various types are available, differing in size of element and speed of response. In infra-red spectrophotometry, where the greatest possible sensitivity is required, it is common practice to use a small receiver, about 2 mm $\times$ 0·2 mm, and to focus on to this a de-magnified image of the monochromator exit slit, using a high-aperture elliptical mirror. The radiation is sometimes chopped at a frequency of about 10 c/s, in order to utilize a.c. methods of amplification and also correct automatically for the zero output of the thermopile. A high-speed thermopile is therefore necessary.

In precise measurements, on the other hand, a high-aperture optical system is often inconvenient, and it is preferable to use a receiving element similar in size to the monochromator slit, and to focus the radiation on it with a spherical mirror of, say, 40 cm focal length, used at approximately 1:1 magnification. The use of chopped radiation is to be avoided, not only because of the difficulty of measuring precisely an a.c. waveform, but also because the high-speed thermopiles required for this method are coated with a minimum amount of black, and tend to be rather selective. Measurements are therefore done by hand, using d.c. potentiometric methods. The zero output of the thermopile is determined in the usual way by means of a shutter, generally placed at the entrance slit of the monochromator. However, owing to the low level of the available radiant power, the zero output must be very stable, and it is often desirable to use a compensated thermopile with two similar receivers.

One instrument which satisfies these requirements and is available commercially has two separate receivers, 9 mm $\times$ 0·5 mm, each made up of three separate elements mounted in line with small gaps in between. The instrument can be supplied with windows of glass, quartz, fluorite, KBr, *etc*, according to the required spectral range. Typical characteristics for each receiver are: resistance 100 ohm; time constant 0·3 sec; sensitivity 28 volts per watt. The thermopile is, of course, unaffected by atmospheric disturbance, and the output fluctuations are almost down to Johnson noise level, which is equivalent to a radiant power of about 10^{-10} watts for a measuring time of one second.

This limit to the performance can only be realized with a highly sensitive electrical indicating system. For d.c. measurements, this usually takes the form of a galvanometer amplifier, *i.e.*, a reflecting galvanometer provided with photoelectric means for detecting the movement of the galvanometer spot. One instrument of this type uses a five ohm primary galvanometer, two balanced selenium cells as the photoelectric detector, and a second spot galvanometer as the output indicating meter. The overall sensitivity is about one metre per microvolt, and the noise level not greatly in excess of the Johnson noise in the primary galvanometer circuit.

The Schwarz linear thermopile should be operated with the image of the monochromator exit slit lying within the confines of the receiver, for if radiation

falls on the reflecting surfaces behind the receiver, the behaviour of the instrument will be irregular. For the same reason, the thermopile is unsuitable for measuring intensity in an extended beam. The sensitivity is not sufficiently uniform over the receiving area to permit precise absolute measurements of radiant power. However, the principal use of this thermopile is in making relative measurements of radiant power at different wavelengths, and the sensitivity is sufficiently uniform for this, provided the distribution of radiant intensity over the receiver does not change excessively with wavelength. Thus, the entrance slit of the monochromator should be illuminated as uniformly as possible, and if the width of the exit slit has to be changed, check readings must be made to allow for a possible change in thermopile sensitivity.

For absolute measurements of monochromatic radiation, the two millimetre circular target thermopile is more suitable than the linear instrument, for although the sensitivity is less, it is also more uniform. In many cases, however, it is unnecessary to measure radiation absolutely over a range of wavelengths. For example, a photocell can be calibrated absolutely by measuring its relative spectral sensitivity with a linear thermopile, and then its absolute sensitivity at a wavelength where sufficient energy is available to permit the use of a more suitable detector.

Measurement of Spectral Selectivity of Thermopiles

A Schwarz linear thermopile is fairly non-selective over the transmission range of its window, but for accurate work the selectivity must be examined. Measurements made with total radiation yield some information, but the ultimate aim is a determination of the spectral selectivity, *i.e.*, the sensitivity as a function of wavelength.

Over the visible region of the spectrum, the appearance of the thermopile receiver usually affords a sufficient guarantee of non-selectivity. The reflection factor of the black is usually less than two per cent, so that significant variations in absorptivity through the spectrum would be evident from the colour of the black.

In the infra-red, the selectivity can be measured by a method similar to that described for total radiation measurements, *i.e.*, the sensitivity is compared, in this case wavelength by wavelength, with that of a black-body thermopile. However, the black-body instrument described in the previous section has rather low sensitivity, and by reason of its circular receiving area, is not well adapted to collecting the radiation from a monochromator slit. A more suitable form of instrument which is used at the N.P.L. is shown in *Figure 4*.

This is of the compensated type, with two similar receivers, each of which is rectangular in shape and provided with two thermocouples. The ratio between the total surface area of the receiver and the

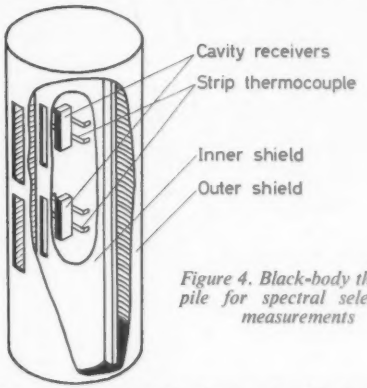


Figure 4. Black-body thermopile for spectral selectivity measurements

receiving area is considerably less for this instrument than for the one previously described, so that the sensitivity is higher. The cavity is correspondingly less black, but auxiliary measurements on the paint with which the interior is coated enable the spectral absorptivity to be predicted to about half a per cent or better. The sensitivity is, of course, still much less than that of a Schwarz thermopile. This is not a serious disadvantage, however, for since the absorption factor of a black does not vary rapidly with wavelength, spectral selectivity measurements can be made with quite wide spectral bandwidths. It has in fact been found possible to measure selectivity at wavelengths ranging from the visible out to about 15 microns.

Spectral selectivity has also been investigated at the N.P.L. by the alternative method of measuring diffuse reflection factors. To avoid the disturbing effect of window reflection, the measurements were made, not on the Schwarz thermopile itself, but on a specially-constructed comparison thermopile, which was operated in air and had a plane, continuous receiving surface.

In measuring diffuse reflectivity, it is desirable to concentrate as much as possible of the reflected radiation on to the measuring instrument, so that the polar reflection curve is properly integrated, and sufficient energy is available for measurement. In principle this can be done by placing the thermopile under test and the measuring instrument in conjugate positions close to the centre of a hemispherical mirror. In practice, however, it is difficult

to get the two instruments close enough together to avoid excessive optical aberration, and the method was therefore modified so that the test thermopile could act as its own measuring instrument. The mirror was arranged to tilt so that the reflected radiation could be alternately directed on to and away from the thermopile; the reflectivity was then determined by measuring the proportional change in thermopile response as the mirror was tilted.

Absolute Radiometry

We turn now to the absolute measurement of radiation, and to the instruments, the so-called absolute radiometers, which are used for this purpose. The subject has a long history, and in the space available it is impossible to do more than illustrate the general principles of measurement. The description will be drawn largely from experience at the N.P.L., and will therefore exclude much important work done elsewhere.

An absolute radiometer is essentially a thermal detector provided with some means for measuring the energy absorbed by the receiver on an absolute instead of merely relative scale. It is usually even less sensitive than an ordinary air operated thermopile, and is therefore only suitable for measuring fairly intense radiation. However, it has already been shown how measurements made with total radiation at high intensity levels may be extended by the intercomparison of progressively more sensitive detectors, to cover the whole range of measurement, with regard to both intensity and spectral composition. It follows that there is considerable latitude of choice as to the conditions in which the absolute radiometer is used, and generally measurements are made with total radiation of intensities of 0·01 W/cm² or more.

It is possible to measure absolutely the heat absorbed by a radiation receiver in terms of known specific or latent heats, using the ordinary methods of calorimetry. Instruments based on this principle have been described, but the method now universally employed in accurate work is one of substitution, in which known electrical power is supplied to the receiver, and its heating effect compared with that of the radiation to be measured.

The Coblentz Radiometer

Figure 5 illustrates the essential features of the radiometer used by Coblentz in an extensive series of measurements on the Stefan–Boltzmann constant. The Coblentz radiometer consists essentially of a blackened, electrically conducting, strip of manganin or platinum foil, behind which is mounted a linear radiation thermopile which responds to the thermal radiation emitted from the back of the strip, and so acts as the temperature-sensitive element. The thermal time constants of the strip receiver and of the thermopile are quite short, so that steady state

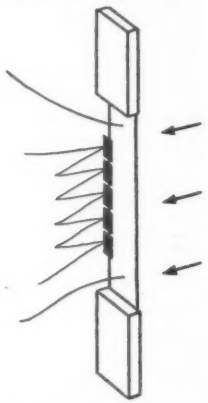


Figure 5. Principle of the Coblentz absolute radiometer

readings may be used, and the procedure of measurement is simply to adjust the electrical power input to the strip until it produces the same response in the thermopile as the radiation being measured. The electrical power is measured using two fine potential leads attached to the strip a short distance from each electrode.

The Coblentz radiometer is intended for measuring radiant intensity in a uniform field of radiation, and it is therefore necessary to know the effective area of the radiometer which relates intensity to measured electrical power. Since the edges of the strip tend to be ragged, they are shielded from radiation by an external slit. The effective area is then precisely defined by the width of this slit and by the separation between the potential leads.

The accuracy of an absolute radiometer is ultimately dependent on the degree to which radiant and electrical heating are equivalent in their effect on the temperature-sensitive element. If the two types of heating are to be equivalent, it is necessary either that the sensitivity of the receiver should be uniform, or that the spatial distribution of electrical heating should be the same as that of the radiant heating, which in general is uniform over the receiving area. In practice neither condition can be fulfilled exactly, and it is therefore desirable to satisfy both as closely as possible so that the resultant error may be of second order magnitude.

In the Coblentz radiometer, the distribution of radiant heating differs from that of electrical heating, since the edges of the strip are screened from radiation, and it is therefore essential that the

sensitivity should be uniform across the strip. Coblentz found that it was, in fact, sufficiently uniform. To some extent, however, this result must be considered fortuitous, for although the temperature-sensitive element averages the temperature rise of the strip, there is no guarantee that the surface heat loss from the strip, by air conduction and convection, is uniform.

The Guild Drift Radiometer

In this instrument, special care has been taken to ensure uniform sensitivity, by making the receiver so thick that lateral heat conduction through it averages out any non-uniformity in the heat losses. The receiver consists of a copper disc of 4·5 cm diameter and 0·6 cm thick, supported by an ebonite ring. The temperature-sensitive element consists of a number of thermocouple wires soldered to the rear surface of the disc, in a pattern which gives good averaging of the temperature rise. The cold junctions are taken, not to a heat sink, but to a similar disc mounted behind the first, so that the radiometer is compensated. A circular diaphragm of slightly less than 4 cm diameter, mounted close to the surface of the receiving disc, defines the effective receiving area.

Electrical heating is supplied to the receiver by means of a resistance element cemented to the front face of the receiver. It consists of a circular disc of thin platinum foil, which is slightly smaller in diameter than the receiver, and is cut to produce a zig-zag conducting path. A pair of fine wires are soldered to each end of the element to serve as the current and potential leads.

The process of measurement is a substitution one, similar to that used with the Coblentz radiometer, except that the thermal time constant of the receiver is too long to permit steady-state readings, and measurements are therefore made on rates of temperature rise and fall, using a timed cycle in which heating periods alternate with equal periods of cooling. When the receiver is heated electrically, some of the heat generated in the current leads is conducted to the receiver; however, the magnitude of this effect can be determined in a separate experiment, by passing current through each current lead and out through the corresponding potential lead. It may be noted that in the Coblentz instrument, the heating effect of the current leads is dealt with by a guard ring technique.

New Types of Absolute Radiometer

The absolute measurements described above must be supplemented by measurements of the reflection factor of the blackened receiver. For this purpose, Coblentz used a hemi-spherical mirror and a thermal detector. With the Guild instrument, absolute measurements are made using visible radiation only, and the reflection factor for this is determined either visually, or by a photoelectric method using an integrating sphere. The measurement of diffuse reflection factors is a troublesome feature of absolute radiometry, and the N.P.L. has therefore developed recently a modified version of the Guild radiometer, in which the receiver is coated with a specularly reflecting black.

This instrument is similar to that described in the section on the measurement of radiant power, in that the receiver consists of a 1 cm diameter anodized aluminium disc, to the rear surface of which is cemented a number of series connected thermocouples. To the front surface of this receiver is cemented a disc of thin mica two centimetres in diameter. The front surface of this carries the heating element, which is in the form of a thin gold film deposited by vacuum evaporation in a grid pattern which covers approximately the same area as the receiver.

The black is deposited on top of the thin film heating element. It consists of a thin film of platinum mixed with tin oxide, prepared by a sputtering process. The reflection factor of this black is seven per cent or less out to a wavelength of two to three microns, and although this factor is somewhat higher than that of a diffusely-reflecting black, it can be measured with greater precision owing to the specular nature of the reflection.

This radiometer is similar with regard to sensitivity and time constant, to the thermopile previously described. Steady state readings may be made with an exposure time of about one minute, and it is therefore unnecessary to resort to the lengthy drift method used with the Guild radiometer.

Another absolute radiometer has been constructed, in which the difficulties of reflection factor measurements are avoided by using a black-body or cavity receiver. Apart from the electrical heating element with which the receiver is provided, this instrument is essentially the same as that described in the section on selectivity measurements, and has a similar response time and sensitivity. As with the last radiometer described, the heating-element consists of a thin gold film on a mica sheet cemented to the base of the receiver. The absorption factor of the cavity receiver for all kinds of total radiation can be estimated to within 0·1 per cent, so that this instrument permits a rapid and direct measurement of absolute radiation intensity, in a wide range of conditions, and with no auxiliary measurements whatever.

POLYURETHANE FOAMS

S. H. MORRELL

Research Association of British Rubber Manufacturers, Shawbury, Shrewsbury

Following a discussion of the chemistry and technology of polyurethane foams, the author describes the properties of some typical members and also their applications. It is concluded that although for many purposes the polyether-urethane foams are preferable, polyesterurethane foams show advantages, particularly where a large absorption of energy is required.

IN RECENT years foamed polyurethane rubber has appeared as a competitor to latex foam in many applications. The purpose of the present article is to describe the rather unusual behaviour of these foams and to show how their unique properties may be employed to the best advantage. The chemistry of the polyurethane foams is similar to that of the solid polyurethane rubbers, described in a previous article by the present author[1] and the polymers may be derived from either polyesters or polyethers; there are important differences between the two materials, both in their technology and in their physical properties, although their general chemistry remains the same.

Chemistry of Polyesterurethane Foams

Waxy polyesters, of molecular weight 2000 to 4000, formed for example by the reaction between ethylene glycol and adipic acid, are reacted with a di*iso*cyanate: this increases the molecular weight to between 4000 and 15,000 and also provides reaction centres which are utilized in the next stage of the process:

$$\text{HO–polyester–OH} + \text{R(NCO)}_2 \longrightarrow \underset{\text{polyester}}{\text{OCN.R.NH.CO.O}} \qquad \underset{\text{polyester}}{\text{O.CO.NH.R.NH.CO.O}}$$

For rigid foams the polyester chains are modified by the incorporation of small amounts of trihydric alcohols, thereby giving further reaction sites along the chains for the cross-linking stage of the reaction. Flexible foams require only a small amount of the chain branching which this produces.

Terminal *iso*cyanate groups may be reacted with water to give carbon dioxide:

$$\begin{matrix}\text{–NCO} \\ \\ \text{–NCO}\end{matrix} + \text{H}_2\text{O} \longrightarrow \begin{matrix}\text{–NH} \\ \diagdown \\ \text{CO} + \text{CO}_2. \\ \diagup \\ \text{–NH}\end{matrix}$$

A similar reaction also occurs between water and some of the excess *iso*cyanate present in the reaction mixture. It is this liberation of carbon dioxide which is the key to this method of producing foamed rubber and which controls the density of the final product. Simultaneously, cross-linking of the chains also takes place. This can occur at any –NH or –OH group in the chain, *e.g.*,

$$\begin{matrix}\text{–R.NH.CO.O–polyester–} \\ \\ \text{–R.NH.CO.O–polyester–}\end{matrix} \qquad \text{R'(NCO)}_2 \longrightarrow \begin{matrix}\text{R.N.CO.O–polyester–} \\ \text{NH.CO} \\ \diagup \\ \text{R'} \\ \diagdown \\ \text{NH.CO} \\ \text{R.N.CO.O–polyester–}\end{matrix}$$

The heat of reaction is considerable and is sufficient to promote the cross-linking reaction and prevent collapse without additional heat. Addition of catalysts may be made to accelerate the rate of production of carbon dioxide as well as to increase the temperature of the foamed mass, which in some cases may exceed 100°C.

It is very important that the rate of production of carbon dioxide should be closely controlled and should bear a definite relation to the final degree of cross-linking required. The cross-linking should take place rapidly enough to trap the gas in small bubbles as it is formed. If the reaction occurs too slowly gas will be lost and a poor foam will result. If on the other hand the reaction is too fast, the gas will cause large bubbles in the foam and the density will be uneven. The final degree of cross-linking achieved depends on the number of *iso*cyanate groups present, which in turn is a function of either the chain length of the original polyester or the frequency of trihydric alcohol groups in it. The higher this degree of cross-linking, the more rigid the foam which is produced. In addition, variations in the rates of the various reactions will also give open or closed cells, or a mixture of the two. Rigid foams contain a high proportion of closed cells, but flexible foams should preferably be of the open cell type to avoid shrinkage.

It is thus seen that here there is a distinction between the production of polyurethane foams and natural latex foaming, where the latex and vulcanizing ingredients are whipped into a foam and then

cured before the foam has time to settle. In the polyurethane the bubbles of gas are more even in size and hence the density is much less variable. Further, only an open cell structure is possible in latex foam.

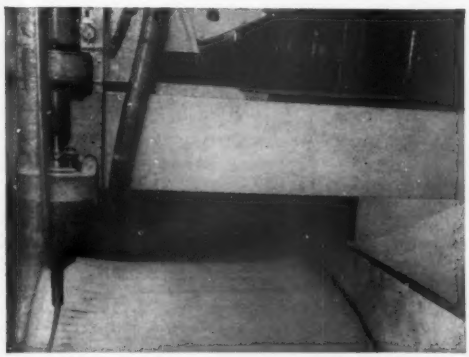


Figure 1. Formation of polyurethane foam (by courtesy of I.C.I. Ltd)

Technology of Polyurethane Foams

In making polyurethane foams by the 'one-shot' process, the polyester or polyether *iso*cyanate, emulsifier, catalyst and water are carefully metered into a traversing, mixing and delivery head and spread immediately in a thin layer over the surface of a slightly inclined tray which in its turn is gradually moved along. Foaming starts immediately and is complete in about two minutes, approximately the time taken to spread the liquid over a full sized mattress. (The formation of polyurethane foam is shown in *Figure 1*.) When the *iso*cyanate used is toxic, the foamed slab is passed into a ventilated compartment to cool, and after about 20 minutes the foam may be handled.

An alternative method involves the initial preparation of a 'pre-polymer' by reacting the polyester or polyether with *iso*cyanate and then feeding this pre-polymer into the mixing head with the emulsifier, catalyst and water. In this way the necessity for metering free *iso*cyanate is avoided and the mixing operation controlled more readily. The balance between the reactions causing foaming and cross-linking is also more easily regulated. A post cure of two or more hours at 120°C is frequently required in order to develop maximum properties. The continuous production of polyurethane foam is shown in *Figure 2*.

All polyurethane foams have a domed upper surface or 'skin' when prepared in this way, and this is removed by means of a band knife (see *Figure 3*). Alternatively, shapes may be cut from the slab by means of red hot wires. Scrap is collected and chopped into small pieces and used as filling for cushions and soft toys, thereby helping to minimize the economic loss which otherwise would occur. A recent development is the use of a spray gun to distribute the mixture of liquids immediately before foaming takes place. A foam backing can thus be applied to rugs, carpets, wood or metal, and pieces of foam may also be invisibly joined by this means.

Properties of Polyurethane Foams

Polyurethane foams are resistant to oxygen and ozone, and may thus be used in exposed positions or at elevated temperatures, although some darkening in colour may occur. They are resistant to oils

Figure 2. Continuous production of polyurethane foam (by courtesy of General Tyre and Rubber Co.)

and solvents, but are attacked by strong acids or alkalis. All polyurethanes are liable to hydrolysis by steam but the polyetherurethanes are more resistant than the polyesterurethanes. The polymer itself is not flameproof but is self-extinguishing; extra fire resistance may be imparted by the addition of self-extinguishing plasticizers such as trichlorethyl

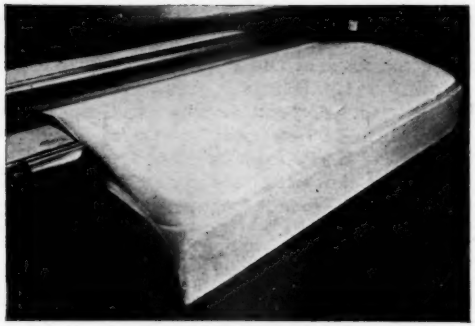


Figure 3. Cutting polyurethane foam with a band knife (by courtesy of I.C.I. Ltd)

phosphate. Natural latex foams, on the other hand, whilst less fire-resistant and inferior at elevated temperatures and in the presence of ozone, oils and solvents, have superior ageing properties in steam, strong acids or alkalis.

The relation between the compressive load and the deflection produced is shown graphically in *Figure 4* which gives a comparison of latex foam, polyether foam and polyester foam. It is seen that with latex foam and polyether foam the deflection increases continuously and smoothly as the load increases. Polyester foam, on the other hand, is relatively insensitive to deflection under light loads, but as the load increases it suddenly ' gives ', a small increase in the load producing a relatively large deflection. The effect may be reduced by mechanically profiling the foam, shaping it by means of hot wires, or coring it, when the load-deflection curve approaches that of a polyether foam. The latter are softer and more resilient than latex or polyester foams, but their load-bearing capacities are not as great. In fact for some purposes, a cushion made partly from polyether foam and partly from polyester foam may give the best compromise of properties.

Hysteresis cycles for natural latex, polyether and polyester foams are given in *Figure 5*, and show that the latter material has the best shock absorbing properties. Prolonged treatment of alternate compression and relaxation gives some softening, but even after a million cycles at a frequency of 240 cycles per minute, polyurethane foam has been shown to return to its original dimensions without appreciable set or other sign of structural breakdown.

The thermal conductivity of polyurethane foams is less than that of cork, slag wool, or expanded PVC, but not quite as low as that of glass wool fibre or expanded ebonite. The density is about half that of cork, expanded ebonite, or natural latex foam, and about the same order as that of slag wool: it is rather more than that of expanded PVC. Both rigid and flexible foams may be made for thermal insulation purposes; the former are

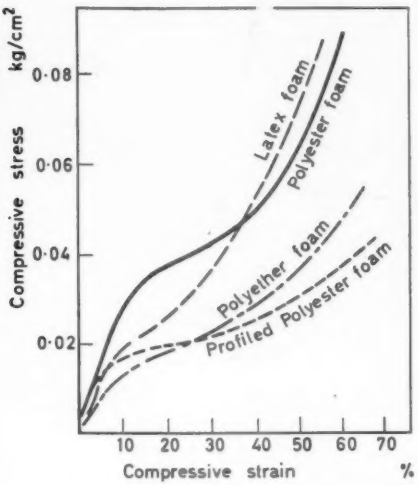


Figure 4. Stress–strain curves for foams in compression (reproduced from Trans. Plastics Inst. *(1959)* **27**, *13)*

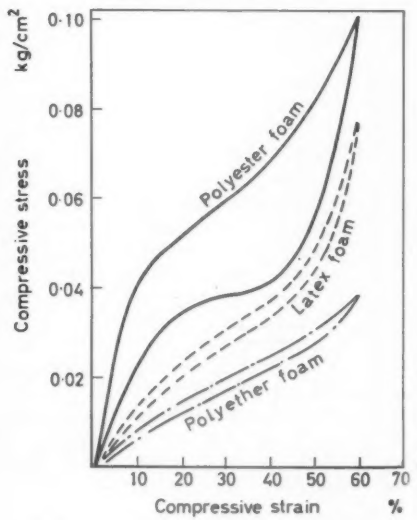


Figure 5. Hysteresis curves for foams in compression (reproduced from Trans. Plastics Inst. *(1959)* **27**, *13)*

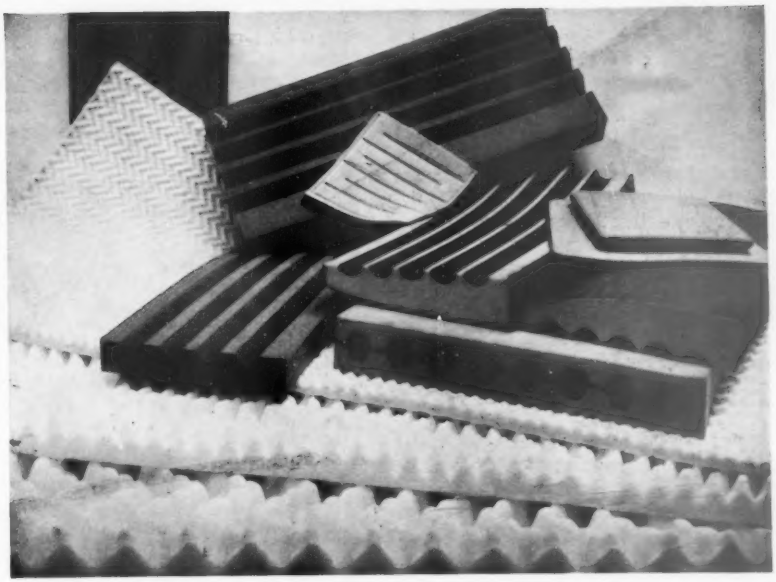


Figure 6. Profiled and cored polyurethane foams (by courtesy of I.C.I. Ltd)

Figure 7. Selection of articles made from polyurethane foam (by courtesy of Dunlop Rubber Co. Ltd

especially easy to stick to most surfaces and may be foamed in place, no extra adhesion being necessary.

Applications

The most important application of polyurethane foams is in cushioning; an estimate of the American market showed that this use of polyurethane foams had increased tenfold in 1958 in comparison with the previous year, and was then 20 per cent of the total usage of polyurethane foams. An elaborate estimate of the possible market by P. B. BAKER[2] suggested that the future lay in foams based on polyethers rather than those based on polyesters, that the cost of these should fall to about the same value as that of foamed latex, but they would have the advantages of resistance to dry cleaning solvents and greater resistance to moisture; they are resistant to mildew and bacteria also.

The shock-absorbing properties of polyurethane foams should make them of interest as packaging materials for bottles and carboys. Packs of foam are used to break the fall when unloading large rolls of paper, beer barrels, and similar heavy objects. Polyurethane foam is unsuitable for machinery as its modulus is too low for use in compression and it is too weak for use in shear, but it is coming into increasing use as a decorative trimming in cars, which at the same time may lessen injury in the case of an accident. Examples of some applications are shown in *Figures 6* and *7*.

The only other current application of polyurethane foams of importance is in pads for garments. Here the tendency for the foam to tear away from the stitches is much less than for latex foam, whilst the resistance to solvents used in dry-cleaning is another asset. In addition, cases have been reported of the spontaneous ignition of latex foam shoulder pads during the dry-cleaning process, presumably due to the ease with which oxygen can attack natural rubber; such spontaneous ignition should not occur with polyurethane foams.

Polyurethane Foams in Commercial Production

The original patents were taken out in Germany in 1941, in connection with the product *Moltopren* (see *Figure 8*). As with the solid polyurethane elastomers, produced by the same firm, the polyesters are marketed under the name *Desmophen* and the *iso*cyanates under the name *Desmodur*. Similarly in America *Multron* and *Mondur* are supplied and in the United Kingdom *Daltocel SF* and *Suprasec SF*. With the raw materials supplied ready for use in this manner, it is often difficult to say whether a particular brand of foam represents a different chemical entity or is manufactured on licence. In the case of *Airfoam SL*, however, the polymer is evidently related to *Chemigum SL*, to which reference is made in the previous paper[1], and which is made by the same firm. Similarly, *Polyfoam* and *Genthane S* are probably related.

Figure 8. Articles made from Moltopren (by courtesy of Fabenfabriken Bayer A.G.)

Conclusions

The number of applications of polyurethane foam is not large at present but without doubt will increase rapidly when the potentialities of this relatively new product have been more fully appreciated by engineers and designers. Its cushioning and shock-absorbing properties, combined with its chemical resistance and resistance to swelling by oils, make it an extremely valuable material. Furthermore, the foam may be manufactured to suit various colour schemes, and is sufficiently attractive in appearance for use without covering. Polyester and polyether urethanes show rather different physical properties, which give one or the other type marked advantages in particular applications.

The author is indebted to an extensive review by J. Acres[3] and to other recent articles[4,5] as well as to those firms who have kindly provided photographs. He is also grateful to colleagues for helpful discussions.

References

[1] MORRELL, S. H. *Research, Lond.* (1959) **12,** 315
[2] BAKER, P. B. *Rubber World* (1958) **138,** 733
[3] ACRES, J. *R.A.B.R.M. Intelligence Division Report* IBC 443
[4] WATTS, J. T. *Rev. gen. Caoutchuc* (1958) **35,** 1571
[5] BUIST, J. M. and LOWE, A. *Trans. Plastics Inst.* (1959) **27,** 13

SEPARATING ISOTOPES OF URANIUM

Development of an Industrial-Scale Process

F. I. HURLEY

United Kingdom Atomic Energy Authority (Development and Engineering Group), Capenhurst, near Chester

When research has demonstrated that a process is feasible, extensive development work is needed to establish the process on an industrial scale. The problems encountered in establishing the gaseous diffusion process for separating uranium isotopes are described to illustrate the nature of this phase of industrial development. Many of the problems arose from the very large size of plant needed to operate this simple process to give the required output—others were introduced by the novelty of the application and the special properties of the basic material. To bring the plant to a state of successful continuous operation demanded close co-operation at all stages between design, development and operating teams, and a concentration of knowledge from many fields of science and technology on the problems involved.

IN DESCRIBING diffusion plants for separating the uranium isotopes stress is usually laid on the large size of the installation. One result of this not unnatural emphasis on size is to forget that the plant is based on a simple physical process, namely that the molecules of the lighter isotope will diffuse more readily through holes whose dimensions are of the order of the mean free path. It is instructive to see what is involved in translating this apparently simple process into a large-scale production plant.

Nature of the Plant

The first thing we must know is the change in concentration of the isotopes that we achieve by passing the gas through a small aperture. An expression derived in the Appendix shows that the change in concentration of the light isotope, Δc, across the aperture is given by

$$\Delta c = c(1-c)\frac{\Delta m}{2m_2}$$

where c is the concentration of the light isotope at the input to the aperture, m_2 is the molecular weight of the heavy isotope, and Δm is the difference in molecular weights of the two isotopes.

The only acceptable compound of uranium for the gaseous diffusion process is uranium hexafluoride, commonly known as hex; it is fortunate that fluorine has only one isotope, so that there is no question of separating the isotopes of fluorine as well as the isotopes of uranium. The molecular weights of the molecules of hex for the two most abundant isotopes of uranium (U^{235} and U^{238}) are 349 and 352. This leads to a numerical value of $\Delta m/2m_2$ of $4{\cdot}3\times10^{-3}$. It follows that the enrichment of the uranium isotopes achieved by a single pass is very small; to obtain a high degree of enrichment of the light isotope, which occurs naturally only to the extent of 0·7 per cent, we must think in terms of several thousand passes. Thus for significant enrichments the plant must inevitably be large and complex, even if the product rate is only small. A diagram of the basic stage unit is given in *Figure 1*; the methods of joining the individual stages of separation together into a continuously operating cascade have been adequately described[1,2] and will not be repeated here.

Many limitations are imposed by the chemical properties of hex. This material is a strong fluorinating agent; it reacts with metals to form metal fluorides, with organic compounds to form fluorinated organic compounds, and with water to form the highly corrosive hydrogen fluoride. After reaction the uranium is no longer in a gaseous form, but becomes a non-volatile fluoride. If any reactions of this sort proceed in the plant the uranium can no longer take part in the separation process. This is particularly serious in the upper parts of the cascade where the uranium is highly enriched; an increase of hold up of this valuable uranium is not acceptable. It follows that the materials of construction of the plant must be carefully chosen and must be stabilized against attack by hex. The amount of attack that can be tolerated is small and hence the range of materials which can be used in the plant is strictly limited. Further, the plant must be designed so that water vapour and other contaminants which would react with hex cannot enter it.

Another major technical problem is the separating membrane. The mean free path of gaseous molecules at atmospheric pressure is of the order of 10^{-5} cm. The aperture for separating the uranium isotopes must be at least as small as this and preferably smaller. That is to say, we must aim to make apertures that are no larger than one hundred times the diameter of most inorganic molecules. Clearly, in order to achieve a sufficient volume flow of gas, the separating membrane must contain a very large number of holes of this order of size. The danger inherent in fine holes is that they will plug and block the passage of gas. It is therefore of paramount importance that the constructional material of the membrane should not react with hex, and the choice of materials is again severely limited. Since the mean free path of the gas molecules increases as the pressure is reduced, the efficiency of separation by the membrane will be greater at lower pressures; and, in fact, much of the plant operates at pressures less than atmospheric.

The importance of preventing inleakage of contaminants is emphasized by these low pressures. If air leaks into the plant it will carry water vapour with it; this will react to produce a solid uranium compound which will block the holes in the membrane. The amount of vapour that can be tolerated is extremely small, and calculations show that the leakage standard must be that of a high-vacuum system. High-vacuum techniques may also be used to remove contaminants adsorbed on the surface of plant components[3,4].

The essential features of a diffusion plant then are that it will consist of a large number of stages containing membranes with many minute holes, that it will be built to a high vacuum standard from materials which do not react with uranium hexafluoride, and that it will be so designed that contaminants cannot enter it at any point.

Design Problems

The design of a plant falls into two stages: flowsheeting, that is to say specifying the quantities and physical conditions required in various parts of the plant to give the output; and detailed design of the components to realise the flowsheet.

The choice of operating pressure is a typical flowsheeting problem. It has already been pointed out that the efficiency of the membranes increases when the pressure is lowered. But the lower the pressure, the larger will be the volume of gas in the plant. A compromise has to be reached between reduced running costs from more efficient operation and increased construction costs from larger components and ducting[5].

In principle, the plant components, with the exception of the separating membrane, are conventional; compressors to circulate the gas, heat exchangers to remove the heat of compression, motors to drive the compressors, valves to control the flows and instruments to provide operating

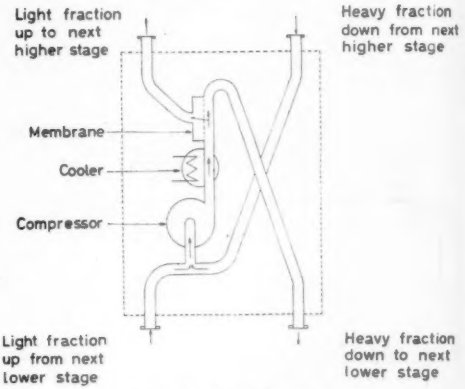

Light fraction up to next higher stage
Heavy fraction down from next higher stage
Membrane
Cooler
Compressor
Light fraction up from next lower stage
Heavy fraction down to next lower stage

Figure 1. Basic stage unit

information. But the conditions of operation are by no means conventional, and the limitations imposed by the restricted range of permissible materials and the very high vacuum standard required have made almost every component of the plant into a new development. The old laboratory high vacuum techniques of sealing wax and plasticine cannot be considered for an engineering plant. Suitable techniques of fabrication and jointing, by gasket and weld, have had to be evolved. The achievement of a plant which, when constructed, will be vacuum tight to a high standard is essentially a matter of following the correct design principles[3].

The plant must operate continuously. There are many plants that run continuously for, say, eleven months in the year and are shut down for maintenance during the twelfth. The diffusion plant as a whole must never be shut down, most components must run for several years without attention and at any time only a small fraction of the components can be off line. A standard of reliability in advance of conventional industrial practice is called for. A large step towards continuous operation is taken by designing a plant of the right materials to a high vacuum standard and ensuring freedom from contaminants. Mechanical reliability of the components is a further essential; valves, compressors, bearings, electric motors must all have a long life. In addition to these components peculiar to the cascade, however, the same standard of reliability in continuous operation must be

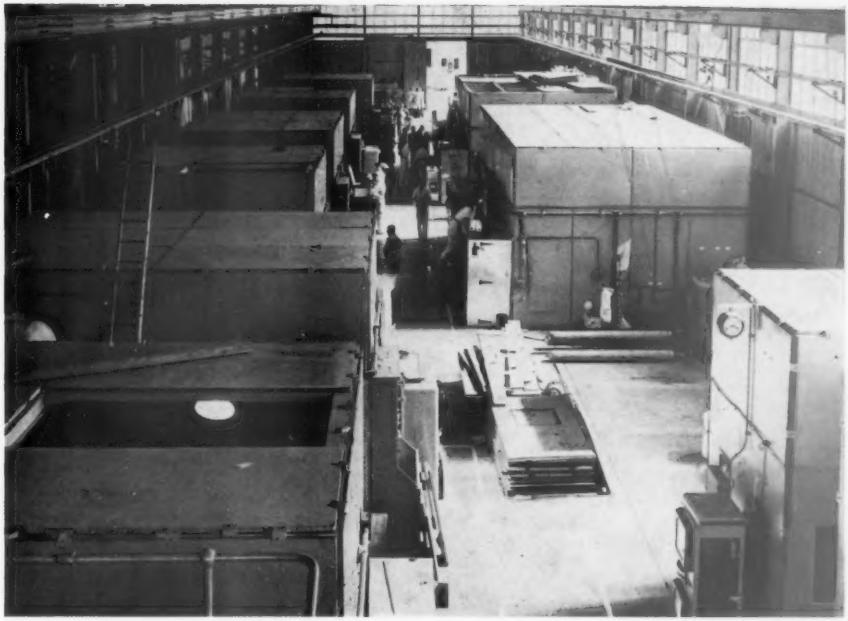


Figure 2. Laboratory for engineering development of plant components

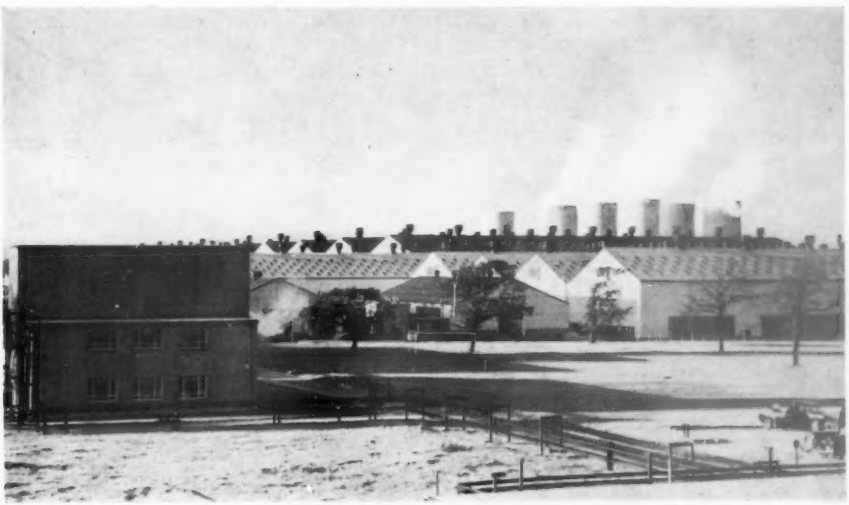


Figure 3. Process buildings, Capenhurst

achieved in all parts of the plant where conventional equipment is involved, such as electrical distribution and water cooling; there is very little advantage in designing components which have a long and satisfactory life themselves if their operation is interrupted frequently by failure of ancillary equipment.

Development Problems

The most obvious requirement for a diffusion plant is the development of a separating membrane. Membrane manufacture has become a specific technology of its own and the production of a satisfactory material has led to many interesting developments in specialized production processes, as well as in more basic physics and chemistry. Several types of membrane have been described, based either on the perforation of a continuous sheet by methods such as etching or anodic oxidation, or on compaction of discrete particles by pressing or sintering[6,7]. Many different techniques have been used to assess membranes; direct measurements of separation have been supplemented by permeability measurements[8] and surface area measurements while x-ray scattering and electron microscopy have been used to study the membrane structure[9].

The designer has needed information on the performance of the membrane for flowsheeting calculations, but he has also called for information on other plant components, and for the development of new components and techniques. The design aspects of high vacuum technique have already been mentioned; the establishment of the design principles, the acceptable fabrication techniques and the optimum types of joint have all been the result of continued development studies. When the plant is built, vacuum testing techniques must be applied to prove that the inleakage is less than the specified maximum. Again, the old laboratory techniques cannot be used because they would be far too time-consuming; vacuum testing methods which can be applied on an industrial scale by semi-skilled personnel have had to be developed[3].

Many engineering developments have been necessary, particularly in connection with the compressors used to circulate hex through the membranes. Both centrifugal and axial compressors have been developed extensively in the aircraft industry, but since hex has a molecular weight some twelve times that of air, its sonic velocity is correspondingly lower and this introduces new problems in compressor design. The diffusion plant is a very large consumer of power, and high efficiency of the compressor is of extreme importance. Another component requiring major technical development has been the seal on the compressor shaft, which isolates the compressor in hex from the shaft bearings running in conventional lubricants. Shaft sealing is a very old engineering problem and many devices have been developed suitable for various conditions. But no existing seal met the diffusion plant requirements of acceptable materials, small leakage of sealing gas and reliability. New principles were explored and several seals satisfactorily meeting the exacting requirements were evolved. *Figure 2* shows the laboratory investigating these problems on full scale components.

The role of the chemist in studying the corrosion behaviour of materials of interest to the designer is obvious. Less obvious has been the necessity to study in considerable detail the chemistry of uranium hexafluoride and related compounds, to aid the designer in specifying flowsheet conditions and stabilization procedures. The reaction of hex with plant impurities as well as with construction materials has been investigated; this is but one example of how the development team has provided a service to plant operators as well as designers.

Establishment of the Process

The problems that confront the operator of a diffusion plant are essentially the same as those met by the operator of any process plant. In their simplest terms, he must get the most out of the plant he has been given, for the least cost. This implies keeping all components operating at optimum conditions and effecting any maintenance rapidly so that as high a proportion of the plant as possible is contributing to the output, and satisfying the conflicting requirement of achieving these aims with a small labour force.

A general view of the diffusion plant buildings at Capenhurst is given in *Figure 3* and a section of the operating area in *Figure 4*. These indicate the large size of the complete plant which is responsible for many of the peculiar operating difficulties. Because it consists of several thousand stages in series it has a long equilibrium time, and it may be several weeks before a change in conditions in the middle of the plant affects the amount of product which is taken from the top. This implies that it is possible for unfavourable conditions to develop and build up over a period of time before they are detected; and also that their effects will continue for some time after the fault is rectified. Deterioration of performance can be prevented by continual watchfulness by the plant operators, but this must be informed watchfulness and a thorough appreciation of plant behaviour is demanded of the operator so that the symptoms can be interpreted correctly.

The size of the plant does not always work against the operator. The existence of many

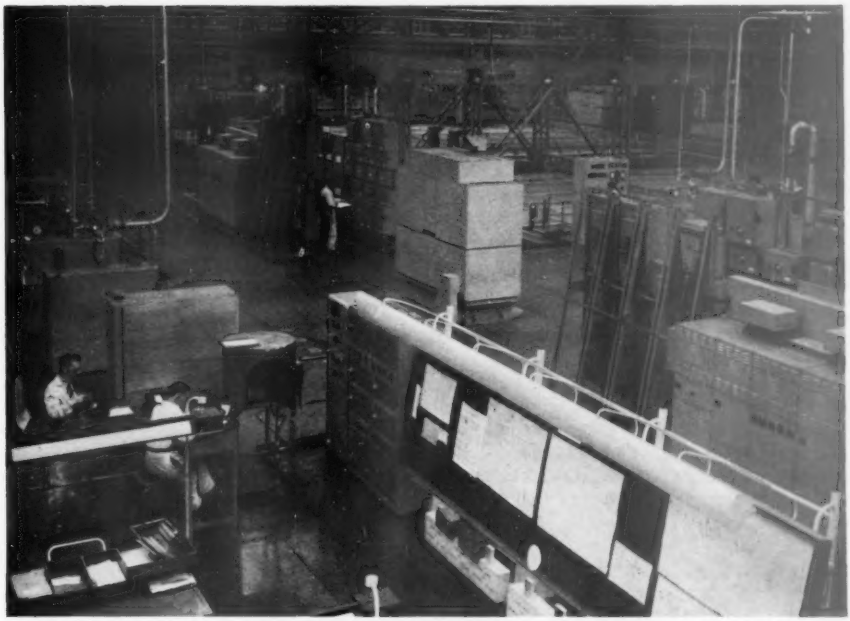


Figure 4. Part of operating area of the Capenhurst plant

Figure 5. Laboratory development of mass spectrometer techniques

identical units means that a statistical approach can be adopted to many problems, and mechanical and electronic methods can be used for handling data. Much consideration has been given to methods of collection and processing of data to provide the operators with the information needed for efficient plant operation.

Requirements for Success

We began by saying that the basic idea of the separation of isotopes by gaseous diffusion is simple. We have seen that, when it is applied to the separation of uranium isotopes on an industrial scale, a large number of technical problems arise and have to be solved before the process can be made to work. The mathematician has had to contribute the theory of cascades, the physicist the membrane, the aerodynamicist the compressors, the engineer novel seals, the chemist a general understanding of fluorine chemistry, and the instrument engineer mass spectrometers and control systems. Development work on mass spectrometers is shown proceeding in *Figure 5*. This large proportion of technical development necessary to realise a simple scientific idea is not unusual. The diffusion plant perhaps provides an extreme example, in that the process is unique and that the peculiar difficulties have made demands of an unusually severe nature in many fields of technology. Because the general pattern is a usual one and this is an extreme example, it is valuable to see what conclusions can be drawn from this particular example. Three points emerge; while none of them is new and each has been recognized before, their importance and general application make it worth setting them down once more.

First, the success of a project of this sort depends on very close collaboration between the design team, the development team and the operating team. The designer must be prepared to discuss his schemes with the development team and must be quick to take advantage of new ideas and techniques that emerge; he must also listen carefully to the demands and opinions of the team who will operate the plant that he is designing. The operating team must be able to visualize the design in operation and to ensure that it is a plant that can be operated efficiently; they must have a sufficient appreciation of the principles of the process to present their problems to the development team and to apply new techniques. The scientists in the development team must guide their work along lines which can be realized in practice, and must develop an understanding of the hard facts of life and of the problems of the designer and the operator. When the plant is operating, development will not be finished, nor will the operating team settle to a dull routine. In addition to dealing with the faults and difficulties which inevitably arise, there is a continuing demand for increased efficiency, either by improved components or modified techniques. If the research team is to make a contribution to the solution of these problems, it must have an appreciation of the process as a whole and of operating techniques and problems, as well as understanding the behaviour of individual components and the properties of plant materials. A less obvious but equally true corollary is that the operating team must contain men who have as deep a technical appreciation of the process as the research and development team.

Secondly, a high standard of engineering practice is essential to success. This applies not only to the general design of major components, but to the detailed design and construction of all parts of the plant. Correct materials must be chosen for all components and the best construction techniques must be used; the importance of these aspects for attaining a high vacuum standard have been mentioned, but they apply generally to other problems. These high standards are necessary, not only to ensure that the plant works, but that it will continue to work reliably over long periods.

Thirdly, the work of the development team must probe deeply into the scientific basis of the problems. Time and again it has been shown that a parochial or *ad hoc* attack on a problem is wasteful and can lead to inadequate answers; the broader the knowledge available, the sooner are difficulties explained and sound satisfactory solutions supplied. It must never be thought that development stops when the plant is built; and here extension of the work of the development group on a broad base assists in the transition from investigations of purely design problems to studies of the operating plant. The development team must include experts in many varied branches of science and technology—we have briefly seen how, for instance, the specialized skills of the theoretical mathematician, the mechanical engineer, the aerodynamicist, the fluorine chemist and the instruments engineer have been employed, but this list by no means completes the full roll of specialists whose work has been necessary to establish a successful plant of the required size and complexity.

This paper is published with the permission of Sir William Cook, Managing Director, and Dr H. Kronberger, Director of Research and Development, of the Development and Engineering Group, United Kingdom Atomic Energy Authority.

Appendix

An indication of the degree of enrichment achieved by passing a gas through an aperture can readily be derived. If m is the mass of a molecule and v its velocity, then, since the energy of all the molecules is the same at any temperature,

$$\tfrac{1}{2}m_1v_1^2 = \tfrac{1}{2}m_2v_2^2 \qquad \ldots.(1)$$

where the subscripts 1 and 2 refer to the light and heavy molecules. If N is the number of molecules per unit volume on one side of an aperture of area A, the number n per unit time passing through can be written

$$n \propto NAv \qquad \ldots.(2)$$

The concentration of light molecules on either side of the aperture is

$$c = \frac{N_1}{N_1 + N_2}, \text{ and } c' = \frac{n_1}{n_1 + n_2}$$

and the fractional change in concentration across the aperture is

$$\frac{\Delta c}{c} = \frac{c' - c}{c} = \left(\frac{n_1}{n_1 + n_2}\right)\left(\frac{N_1 + N_2}{N_1}\right) - 1$$

Substituting from *Equations 1* and *2* this reduces to

$$\frac{\Delta c}{c} = \frac{N_2\,(1 - \sqrt{m_1/m_2})}{N_1 + N_2\sqrt{m_1/m_2}}$$

Writing $m_2 = m_1 + \Delta m$, expanding, and neglecting all terms smaller than $\Delta m/2m_1$ and $\Delta m/2m_2$ in comparison with unity we have

$$\Delta c = c(1 - c)\,\Delta m/2m_2$$

References

1. JAY, K. E. B. *Britain's Atomic Factories* London: H.M.S.O., 1954
2. SMYTH, H. DE W. *Atomic Energy for Military Purposes* Princeton: University Press, 1948
3. KRONBERGER, H. *Proc. Instn mech. Engrs* (1958) **172**, 113
4. ROUBIEX, G. *Second United Nations International Conference on the Peaceful Uses of Atomic Energy*, Paper A/Conf. 15/P/1269 (1958)
5. This and other flowsheet problems are described by: BILOUS, O. *ibid.* Paper A/Conf. 15/P/1264 (1958)
6. MÅRTENSSEN, M. and HOLMBERG, K. E. *ibid.* Paper A/Conf. 15/P/181 (1958)
7. FREJACQUES, C. *et al. ibid.* Paper A/Conf. 15/P/1262 (1958)
8. MASSIGNON, D. *ibid.*, Paper A/Conf. 15/P/1266 (1958)
9. CHARPIN, J., PLURIEN, P. and MOMMEJAC, S. *ibid.* Paper A/Conf. 15/P/1265 (1958)

The Use of Ceramic Materials for Gas Turbine Blades

by J. M. Puri

An article under this title was published in July (*Research* (1959) **XII**, 246). This was one of the entries in the 1958 Waverley Gold Medal Essay Competition. The Editor wishes to express his apologies to Dr T. H. Blakeley of The Morgan Crucible Company, Limited, and Mr R. F. Darling of PAMETRADA (Parsons and Marine Engineering Turbine Research and Development Association) since it appears that much of this article has been transcribed from a paper by these authors previously published in the Transactions of the North East Coast Institution of Engineers and Shipbuilders.

THE BIOLOGICAL ACTION OF ATMOSPHERIC IONS

JOHN RHEINSTEIN

Electronics Institute, Technische Hochschule, Munich

After a brief account of the nature of atmospheric ions, the author describes work carried out to study the effects of positive and negative ions on biological systems, ways in which they are artificially generated and applied, and their affect on human reaction time.

SHORTLY after the discovery in 1752 by Franklin and others that lightning is an electrical phenomenon, the possibility of a biological influence of atmospheric electricity was discussed and investigated. As early as 1777 the Academy of Lyon announced a prize for the best paper concerning the illnesses which are caused by a deficiency or excess of electrical ' fluid ' in the body and the means of curing such illnesses. Much of the early work was conducted with great ingenuity but usually with little understanding of the medical or physical problems involved. During this period many types of electrical therapy were brought into use, but there is much doubt concerning their actual value. Unfortunately the questionable use of electrical therapy, without proper medical supervision is not yet dead.

One of the main problems was a lack of understanding of the nature of atmospheric electricity, which was not explained until later. In 1785 Coulomb showed that the atmosphere is a conductor, though not a good one; however he did not understand the nature of this conduction. Not until 1899 did Elster and Geitel explain the conductivity of the atmosphere and show that there exist positive and negative carriers of electricity, which they called ions.

Nature of Atmospheric Ions

These atmospheric ions are not the same as ions formed by gaseous discharge or by electrolytic decomposition and which are responsible for conduction in gases, at lower pressures, or in electrolytes. Rather atmospheric ions are groups, or clusters, of molecules, the entire cluster having an excess positive or negative charge; that is the cluster has gained or lost an electron. Depending upon the number of molecules in a cluster atmospheric ions are termed as small with three to twelve molecules per cluster, intermediate with 10^2 to 10^3 molecules per cluster, or large with up to 10^6 molecules per cluster.

Due to various factors, particularly ionizing radiation or particles in the atmosphere, a molecule becomes ionized and within a very short period of time, of the order of microseconds, other molecules become electrostatically attracted to the ionized molecule to form the small ion cluster. Intermediate and large ions are formed when a small ion becomes attached to a condensation nucleus. To differentiate between the various types of atmospheric ions one generally does not determine the size directly, but rather the mobility, which is the drift velocity the ion will achieve in an electric field of one volt/cm. The mobility of small ions is between one and two, of intermediate ions between 10^{-3} and 10^{-1}, and of large ions less than 10^{-3} cm/sec/volt/cm. The mobility of small negative ions is about 30 per cent greater than that of small positive ions.

The method now generally employed to measure the ion density of the air is the aspiration condenser method illustrated in *Figure 1*. Air is drawn by a fan between the two plates of a large condenser whereby all air particles, including the ions, acquire a velocity component in the y direction. Due to the electric field E between the plates the ions also acquire a drift velocity component in the x direction. If their mobility is large enough ions of the correct polarity will strike and discharge on the lower electrode allowing the ion current to be measured by means of a micro-microammeter, and thus the ion density may be determined. By varying the length of the condenser plates, the electric field

between the plates, or the airstream velocity the lowest mobility which will still be measured may be controlled. In order to avoid spurious readings due to electrical field disturbances the design and construction of the condenser system is quite critical.

It is now believed that the small and possibly the lighter intermediate ions, due to their greater mobility, have the greatest biological effect. The density of the small atmospheric ions is greatly dependent on the number of condensation nuclei present, and thus it is usually lower, between 50 and 500 ions/cm^3, in heavily populated or industrial areas than in areas where the air is clearer such as in mountains, where the small ion density is between 50 and 1500 ions/cm^3. The effect of condensation nuclei on the small ion density may be easily seen by measuring the sudden decrease in small ion density due to the smoke of one cigarette. In a closed room it takes several hours for the small ion density to rise again to the normal value. Daily and yearly variations in the small ion density have been noted, as well as other variations due to various meteorological effects.

At present little is known concerning the chemical composition of the small ions; however A. P. KRUEGER and R. F. SMITH[1] have shown that, at least in certain instances, a minimum amount of oxygen in the atmosphere (here an artificial atmosphere) is necessary for small negative ions to exhibit their usual effects. Similarly a minimum amount of carbon dioxide is necessary for small positive ions to produce the effects noted in their presence. Thus it seems probable that small ions, at least those which produce the given biological effect, are partially composed of O_2 or CO_2. This is also in general agreement with the known ionization potentials.

Earlier Modern Work

Noting that some climatological phenomena such as mountain sickness, the effect of the Foehn, the influence of mountain and sea climate on the body, the sensitivity of rheumatic subjects to changes in the weather, *etc* could not be explained simply by the well known climatological factors such as atmospheric pressure, temperature, humidity, *etc*, many workers in the early twentieth century postulated an additional factor, namely atmospheric ions. Among these was F. DESSAUER[2], about 1930, who appears to be the first to have conducted reasonably well-controlled experiments with artificially generated ions. For this purpose he employed very fine charged magnesium oxide dust, corresponding to light intermediate ions. Then, under clinical conditions, patients breathed in these artificially generated ions. Dessauer generally observed the effects shown in *Table 1*, many of which were then confirmed, using similar equipment, by Edström.

Table 1. Effects found by Dessauer attributable to breathing atmospheric ions

Positive ions	*Negative ions*
Higher blood pressure	Lower blood pressure
Faster, shallower breathing	Slower, deeper breathing
Higher pulse rate	Lower pulse rate
More errors in remembering numbers flashed on a screen	Less errors in remembering numbers flashed on a screen
General stuffy feeling, headache, nausea	General crisp, fresh feeling

As can be seen from *Table 1* in general negative ions seem to have a desirable effect while positive ions seem to have an undesirable effect. This has generally, though not in all cases, now been verified. It is of interest to note that normal electric heaters, and also some other types of heaters, generate large quantities of positive ions.

In the period before the half century others noted various, often conflicting, effects attributable to atmospheric ions. Effects on the larvae of fruit flies, growth and laying ability of chickens, growth of cancerous tumours and many others were found. Chizevskii, as a result of work which he and others had done, concluded that the presence of atmospheric ions was necessary in order to maintain the vital process of animals.

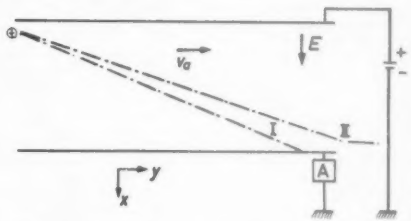


Figure 1. Schematic diagram of aspiration condenser type ion counter. E *is the electric field,* v_a *the airstream velocity,* A *a recording micro-microammeter,* I *the path of a positive ion whose mobility is high enough, and* II *the path of an ion whose mobility is not high enough for the ion to be counted*

Although more recent workers have generally verified the work of Dessauer, and of some of the others, the manner in which many of these earlier experiments were conducted, particularly the manner employed to generate ions artificially, is quite suspect. One of the main problems in the generation

of artificial ions is to insure that no disturbing factors are introduced such as the magnesium oxide of Dessauer, or such other factors as ozone, the oxides of nitrogen or unwanted radioactivity. The relatively recent development of better methods of

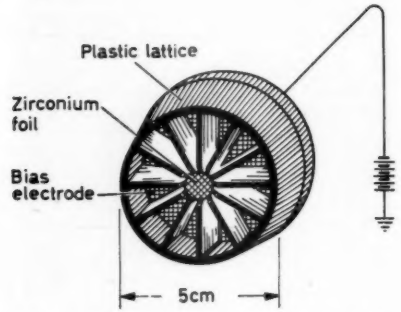


Figure 2. Tritium ion generator, showing zirconium foil in which H^3 *has been absorbed, biasing electrode, and plastic protective lattice. With negative bias voltage, as shown, negative ions are emitted (after* Krueger *and* Smith)

generating ions, and the use of well designed experiments employing proper statistical methods, has made it possible to study further, and with assurance that the results are valid.

The Artificial Generation of Atmospheric Ions

Ions may be artificially generated by means of radioactivity, electrical discharge, ultra-violet and other high energy radiation, heat, and friction, for example splashing water. Generally either radioactivity or electrical discharge is used to produce controlled artificial ionization.

The units employed by the author* to generate ions consist of a small disc, about five centimetres in diameter as shown in *Figure 2*; in the centre of this disc is a smaller, one centimetre disc of zirconium in which 50 mc of radioactive tritium (H^3) have been absorbed. The tritium emits beta particles (electrons) with an energy of 18 keV. These electrons lose their energy very rapidly in passage through the air by colliding with and ionizing air molecules. The range of the electrons in air is approximately 0·6 cm and there is a plastic protective lattice about 1 cm thick in front of the radioactive source so that there is no danger from the radioactivity. The absence of any effects or danger due to the radioactivity, if such a lattice is used has also been verified by actual measurements.

* Kindly furnished by Mr W. Hicks, President of the Wesix Electric Heater Co., San Francisco, California.

Behind the zirconium disc is a large disc electrode upon which a biasing voltage, usually about 900 volts, of either polarity may be impressed. By proper choice of this bias voltage ions of the desired polarity are electrostatically repelled and thus emitted, while those of the undesired polarity are attracted to the electrode and discharged there. The ion current, of the desired polarity, available from this ion generator is about 2×10^9 ions per second. For a subject seated near the generator this corresponds to an ion density of about 10^6 ions/cm^3. This method of generating ions has the advantages of simplicity, portability, and safety.

Recent Work

In recent years several workers have investigated some aspects of the biological influence of artificially generated ions quite thoroughly. One approach, which quite possibly may lead to a widening study in this field, has been the clinical work of I. H. Kornblueh[3]. Patients suffering from hay fever or asthma were exposed to artificially generated negative ions and to a large extent were afforded relief, although in many cases this was only temporary. If, as appears probable, further success is achieved in this line the work would be very fruitful. It is of interest to note that at least one air conditioner now sold in the United States is equipped with a negative ion generator.

Krueger and Smith[1] have, for several years, been studying the effects of atmospheric ions on the trachea. Their first experiments were conducted on tracheal strips of the rat, rabbit, or guinea pig. It was found that the rate of ciliary beat as well as the mucus flow rate was influenced by artificially generated atmospheric ions. Negative ions increased, while positive ions decreased, the ciliary beat and the mucus flow rates. In the case of the

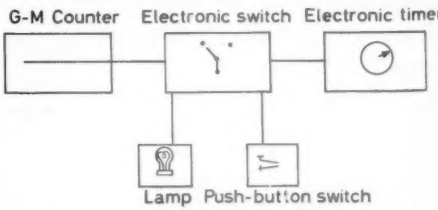


Figure 3. Schematic diagram of method employed for measuring the 'push button' reaction time

ciliary beat the change was approximately 25 per cent from the normal in each direction. Further experiments were conducted on live animals, administering the ions through a tracheotomy, and on unoperated animals. Then experiments were

Figure 4. Experimental arrangement employed. A radioactive type ion generator (polonium source) with blower is shown at lower left. Aspiration type ion counters are shown to each side of subject's head. Units shown are now considered obsolete

conducted on primates (monkeys) and finally on human trachea, in the latter case on tracheal strips obtained at autopsy. In each instance the above response to ions was observed, except that on humans it was impossible to investigate fully the mucus flow rate. As the function of the cilia and mucus of the trachea is to sweep out foreign bodies, a possible correlation between this work and the work of Kornblueh may be made.

Another line which may prove to give valuable results in the near future is a continuation of the work of G. B. NIELSON and H. A. HARPER[4]. They showed that exposure of rats to positive ions produce a statistically significant rise in the succinoxidase activity of the adrenal glands.

That atmospheric ions have effects on the action of enzymes has also been shown by others, and it seems possible that the mechanism by which atmospheric ions exhibit effects on the living body is through the action of enzymes. Some of the effects found by earlier workers, such as the retarded growth of cancerous tumours when exposed to negative ions, and various effects on the respiratory system could possibly also be explained through the action of enzymes.

In addition to the above mentioned work significant effects due to artificially generated atmospheric ions have been observed with regard to the growth of cellular cultures, the carbon dioxide combining power and the pH of the blood, the blood pressure, the adrenal response, and the weight of animals.

Effects of Ions on Human Reaction Time

M. KNOLL, G. F. LEONARD, and P. E. HIGHBERG[5] made a preliminary investigation concerning the effects of artificially generated atmospheric ions on the 'push-button' reaction time of humans, and an effect was observed. The principle of measurement of the 'push-button' reaction time is shown in *Figure 3*. Here random pulses, generated by means of a Geiger–Mueller counter triggered by cosmic rays or another radioactive source, start an electronic timing circuit and cause a lamp to flash. Upon seeing the lamp flash the subject voluntarily reacts by means of a push button switch. This switch then stops the timer, and the time between light flash and the subject's reaction, measured by the electronic counter, is termed the reaction time. The general experimental arrangement used is shown in *Figure 4*. The type of results generally obtained are indicated in *Figure 5*, taken from later data. Although an effect was always observed the direction of the effect, for example whether negative ions would increase or decrease the reaction time, was not reproducible, even for the same person at different times. In these experiments the density of artificially generated ions was about 10^3 ions/cm^3.

For the past year and a half the author* has conducted somewhat similar experiments using equipment based on the same principle. Much higher ion densities were employed, corresponding to

* At the Institute for Technical Electronics of the Technische Hochschule, Munich (Director: Prof. Dr M. KNOLL). This work was done under a contract with the U.S. Dept. of the Army, through its European Research Office. (Contract DA-91-591-EUC-1035 OI-1198-59).

about 10^6 ions/cm^3, and the state of ionization was switched automatically, without the knowledge of the subject or the experimenter, about every three minutes. The data for *Figure 5* were taken from these experiments. The distribution curves of the results of about 150 experiments, involving over 9000 measurements, are shown in *Figure 6*. In order to be able to compare the results from different experiments the ordinate is the percentage deviation, in any one experiment, from the mean of the reaction times recorded when no ions were being artificially generated. As may be seen from the figure it is probable that the distribution of reaction times obtained when the subject was breathing ions is significantly different from that obtained when the subject was not breathing artificially generated ions. This has also been statistically verified, the difference between the distributions being significant at a level of significance of less than 0·0005. That is the probability that the difference between the distributions is *not* due to random variation is greater than 99·95 per cent.

The non-reproducibility mentioned above was also found in these experiments, but further work has begun to clear this up. In experiments with artificially generated positive ions it has been found that a very great ion density, breathed in through the mouth, causes, in about 90 per cent of the cases, a lengthening while a much smaller density usually brings about a shortening of the reaction time. Similar experiments with negative ions are now in progress.

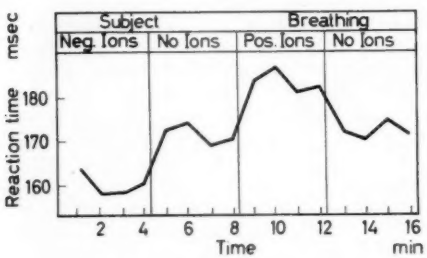


Figure 5. Typical results of a single experiment showing the effect of atmospheric ions on human reaction time. Each point is the average of five measurements

Conclusion

Definite, statistically significant effects due to artificially generated atmospheric ions have now been demonstrated. As is usual the more one investigates a subject the more questions arise, at least at the beginning. At present there are many important questions still unanswered. For example the question of the length and degree of exposure for optimum results is quite important. Static exposure to high ion densities for long periods of time seems generally to have a detrimental effect. Thus a changing exposure, as found in nature, may be required both for the general well being and for the display of therapeutic effects. In addition to the above effects it is probable that at least a part of the 'freshness' of the air in some areas, which modern air conditioning has not yet been able to reproduce, is due to atmospheric ions. It also appears likely that there are other factors both physiological and psychological, which act in conjunction with atmospheric ions.

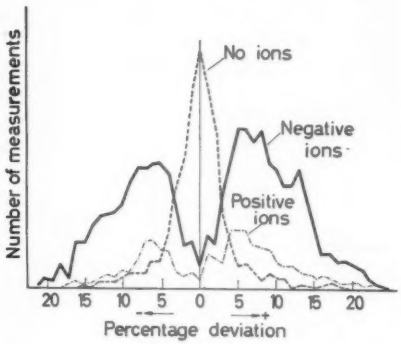


Figure 6. Distribution of reaction times, with and without artificially generated ions. The ordinate is the deviation, in any one experiment, of the measured values from the mean value obtained when no artificially generated ions were being inhaled by the subject

Up to the present relatively little work has been done in this field, but with the increasing growth of the field of medical electronics it is hoped that further work, not only on a phenomenological basis but also on the theory of such effects, will be accomplished.

References

[1] KRUEGER, A. P. and SMITH, R. F. *Proc. Soc. exp. Biol. Med.* (1958) **98**, 412
[2] DESSAUER, F. *Zehn Jahre Forschung auf dem Physikalisch Medizinischen Grenzgebiet* Leipzig, (Thieme Verlag, 1930)
[3] KORNBLUEH, I. H. and GRIFFIN, J. E. *Amer. J. phys. Med.* (1955) **34**,
[4] NIELSON, C. B. and HARPER, H. A. *Proc. Soc. exp. Biol. Med.* (1954) **86**, 753
[5] KNOLL, M., LEONARD, G. F. and HIGHBERG, P. E. *Report of the Dept. of Electrical Engineering* (Princeton University (1958) unpublished)

THE FUTURE OF INLAND FREIGHT TRANSPORT

G. F. FIENNES

Line Traffic Manager, British Transport Commission, London, N.W.1

Bearing in mind the increasing demand for inland freight transport, the author discusses the technical and economic advantages of road, rail and water transport and their future development

THE FUTURE of inland transport depends partly on techniques, partly on economics, and on politics a great deal. Whatever the temptations, however, neither politics nor general economics are part of my task. There remains the technical study. Firstly, the demand for transport: quantity and direction. Secondly, the choice of transport: rail; road; air; waterway; pipeline; transmission line. Thirdly, the present technical advantages and disadvantages of each form and fourthly, the technical advances in prospect for each and in particular for railways.

Demand for Transport

The demand for freight transport like the habit of travel is growing. The average expectation is that national production will increase by about 60 per cent in the next ten or fifteen years. Of that increase over one half will be translated into a demand for transport.

The geographical demand for transport is unlikely to alter much. There are powerful opposing forces in a state of approximate balance. The social interest requires work to be brought to people rather than people to their work. When a basic industry in any area declines, the Government encourage other industries to move in. The object is a natural one, to diversify the work in each area and to spread the social risk. Tyneside, Lancashire and South Wales are cases in point.

On the other hand the growth of the industries depending on import in bulk has stimulated development of the sea-board. The Tees, the Thames, the Humber, Southampton Water, Milford Haven have within the last generation attracted the great oil companies, Fisons, Fords, I.C.I., and many more (see *Figure 1*). On the whole this force has been contained. It has not and will not often demand a radical alteration in the direction of transport or in its form.

The Choice of Transport

Of what sort shall transport be? The national interest demands that transport shall be efficient in the service of the nation; streamlined, taut, spare, always down to its fighting weight; fighting for the best transits and lowest costs for our exports and our consumption at home. The broad principles are convenience, reliability, speed, safety and economy.

The order of importance of the principles varies enormously. Many industries are already geared to supply at a few hours notice; fruit, flowers, vegetables, fish, meat, newspapers, come from the source of supply through, maybe, both wholesaler and retailer to consumer in that time. More and more industries producing and distributing consumer goods have come to recognize that by gearing themselves in the same way they reduce the stocks which they hold at their premises and in the pipeline. One very large concern by taking thought released last year in this way £20 million for other capital use. To these industries the emphasis is on reliability and speed.

At the other extreme are the basic industries of coal, oil, iron and steel. One of their principal interests is that their processes shall not be held up by a hiatus in the arrival of their raw materials or by congestion in the despatch of their finished products. They demand reliability but it means something very different from reliability in, say, the

fish trade. Speed is less important; a steady flow more. Economy, since transport is a larger component of the final cost of the product, is vital. Between these extremes is an infinite range of shades of demand.

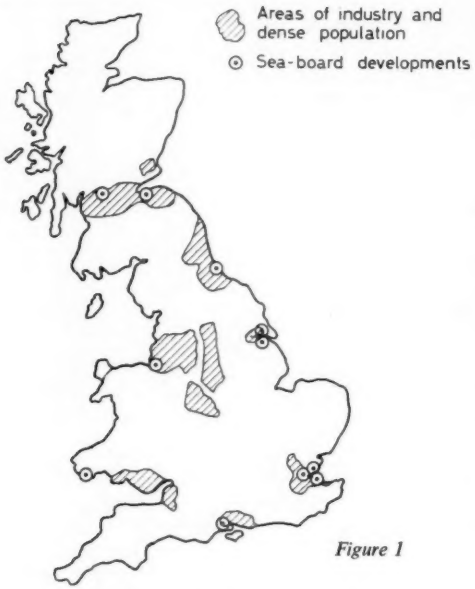


Figure 1

Relative Technical Advantages

The next step is to examine the relative advantages of the forms of transport in common use—at present road, rail, water.

Water in this context means inland rivers and canals. They have advantages in convenience and economy especially for the industries with long traditions which began their development at the waterside. For them discharge of coal or oil or goods in bulk from barge employs existing capital. The lighter, like the railway wagon, but unlike the powered barge or the lorry, awaits discharge without complaint and with little capital standing idle. In 1957 industry preferred inland water transport to the extent of rather less than one per cent of the carryings by rail.

The advantages of road transport are massive. For convenience a flexible unit; available like the railway train at all hours of the day or night but ready to move at any moment unhampered by the organization required to match the wagons with an engine, a crew, a brake van, and a path over the line. The unit is small; more closely matched to the unit of the country's trade than is the train of 500 tons. Lorry does not wait for lorry as wagon must wait for wagon. There are opportunities for individual service in the time of loading and unloading which railways, running to a time table, do not offer.

In regularity the advantage of the smaller unit direct from door-to-door under the personal control of its own driver, predominates over the railway's practice of transhipment between road to rail and of combination with other wagons to form a train —on rail all this controlled by a series of individuals, none of whom has a direct interest in the journey as a whole.

In speed an average of twenty miles an hour allows overnight transits up to, say, two hundred and fifty miles. There are few large centres of production or consumption in the country which are farther than that distance apart. The speed of the lorry, which by no means matches that of the railway train, is firstly good enough and secondly is a true overall door-to-door speed, whereas the speed of the train is often of less account by reason of the delays inherent in the road/rail transfer and in marshalling yards.

In safety the personal supervision of the driver gives road an advantage. The door-to-door service with no transhipment to cause damage or loss, or shunting to cause damage. On the other side of the account are the road accidents costing £175 million a year.

This brings us to the advantage of rail. In convenience the service, nation-wide, ranks high. Next, the railway wagon will stand silently and comparatively cheaply awaiting the senders' or consignees' pleasure. The length of the queue for unloading matters less on rail than on road whether viewed nationally or individually. Thirdly, most of the important industries are expensively geared to rail for inward and outward traffic with private sidings mechanical loading and unloading. Some 95 per cent of railway tonnage is loaded and 74 per cent unloaded in private sidings. Much of the rest—domestic coal for instance—is unloaded in railway yards where the merchants have space for storage.

In regularity the only point to be made for rail at the moment is that express freight trains on the move run reasonably punctually. In winter some 60 per cent are within thirty minutes of time; in summer some 80 per cent The proportion of express freight to the whole is rather more than a third and going up rapidly. For the speed of actual movement rail is far superior. The fish and meat trains at average speeds of over 45 miles an hour; the express freights at 35 and upwards have no parallel on road.

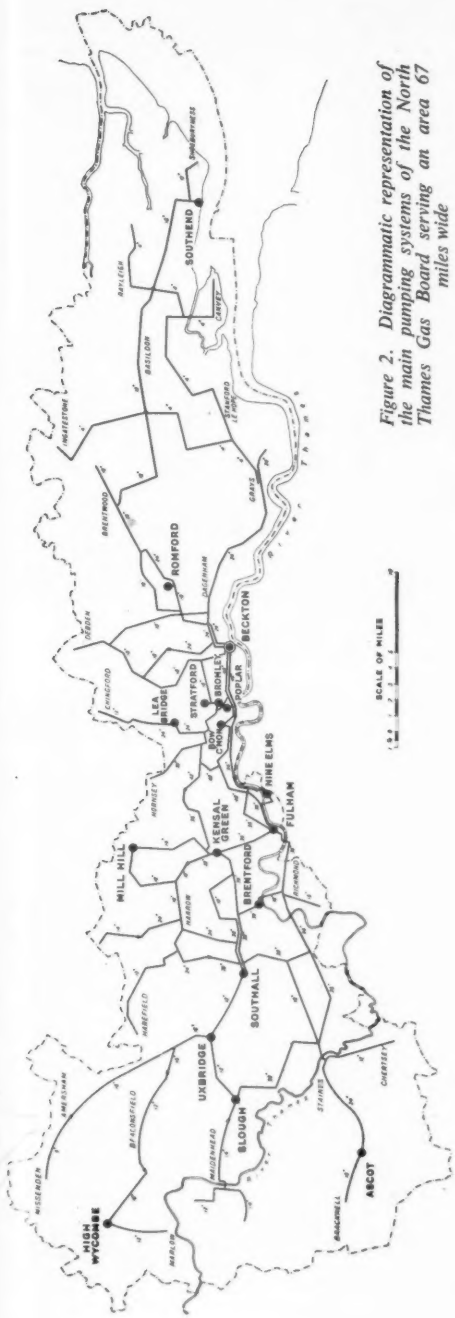


Figure 2. Diagrammatic representation of the main pumping systems of the North Thames Gas Board serving an area 67 miles wide

Technical Advances in Prospect

The survey of the present has included little about transport by water and nothing about air, pipelines, and transmission lines which at present make no significant contribution. We come now to technical advances in each form. For water there appears to be little in prospect. There will be an improvement in the condition and capacity of the principal routes and small advances in the design of craft but nothing that will enable our inland waterways to play a significantly greater part.

The design of aircraft is going ahead by leaps and bounds. For inland transport the development of supersonic or subsonic jets is unlikely to have any relevance. The distances are too short to permit economic operation. Conventional types continue to have the disadvantage common with rail of transhipment at the terminal. They will, no doubt, cream some traffic, both in public service and in private. As with the C licence road vehicle, the advantage of their own transport on tap will prevail on some firms to use their own aircraft. Development is, however, likely to be along lines which will avoid the delays and dangers of transhipment at airports. This development can take the form either of the use of helicopters or other vertical take-off aircraft which will use the senders' and consignees' premises, or alternatively of conventional aircraft capable of absorbing vehicles or containers in its belly.

Transport by air in these forms will have much the same convenience, reliability and safety as the road unit. In speed it will be superior which will not matter very greatly since except over the longer distances road can, and rail will, do everything which is necessary overnight. In cost it will be high. It will be used in two ways; the first to cream some highly profitable traffics largely from road, traffics where the cost of transport is not a significant part of the selling price of the article: secondly, and more sensibly to carry over long distances traffics which are highly perishable. Fish, meat, flowers, newspapers, come to mind.

Apart from road, rail, and air, there are specialized forms of transport for certain products. Transmission lines for electricity; pipelines for gas and oil. Each of these forms has its special attractions. To transport electricity by transmission line over long distances enables coalfired power stations to be located in the coalfields, oilfired near refineries or ports, nuclear on remote wastes. The losses of energy in transmission, however, are severe. And the economics are not so heavily on the side of this development that it has gone very far. In the very long run, of course, the tendency to nuclear power

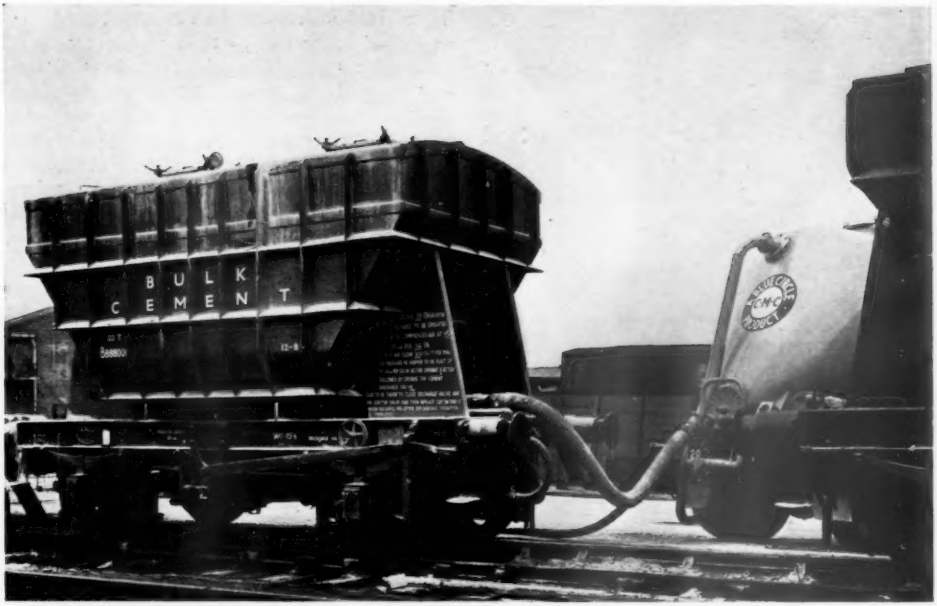


Figure 3. 'Presflo' wagon containing bulk cement showing discharge by compressed air to road delivery vehicle

stations will increase the length of transmission but not in the period under study.

On the other hand, pipelines are already contributing and will develop for the carriage of gas and oil in bulk (see for example, *Figure 2*). For gas the abolition of small and inefficient gas works and the supply of gas from neighbouring large towns worthy of large and efficient plants is already commonplace. For crude oil they are, of course, the normal form of transport in the principal oil-producing countries. They exist in England as a relic of the war; Pluto under the Channel and the internal network of lines for the Royal Air Force. The applications were different; mainly crude oil abroad, mainly petrol here. Technically both are possible. The only form of oil which presents great difficulty owing to viscosity is fuel oil.

Development of pipelines here for refined oil is unlikely to go very far. In a country so densely populated and so trigger happy at the approach of the invader of private right the time and energy needed to secure wayleaves will daunt the hardiest oil company. To lay lines above ground along railways or roadways, raises obvious risks of fracture and fire in the course of a rail or road accident. To bury the lines increases the cost and problems of installation, but some development is in prospect.

On the other hand there is a powerful force working in favour of pipelines for crude oil. Imports in deep-sea oil tankers of over, say, fifty thousand tons can come in at none of the regular ports of entry—Thames, Humber, Mersey, Southampton Water, Bristol Channel. Consequently Milford Haven and other inlets remote from the present refineries are being developed or prospected. The oil companies then have the choice of taking the refinery to the new port of entry or of transporting the crude oil from the new port to the existing refinery. Since most of our refineries are virtually new and highly efficient the choice seems to be to transport the crude oil. The economics of this transport are in favour of pipeline. One line from the Gareloch to Grangemouth has been in use for many years; a line from Milford Haven to Llandarcy is being laid; one from Holyhead to the Mersey is under study; and so on. But apart from the transport of crude oil over short distances in the less inhabited parts of the country, development of pipelines is unlikely for the reasons given.

The technical advances in prospect for road transport are in two parts, the track and the vehicle. There is a network of motorways planned along the principal arteries of the country. There is a great programme of improvements to the existing

trunk roads, in all 1500 miles. The vehicle itself can develop little in capacity owing to the physical limitations of the track, especially of the bridges. It is unlikely to develop greatly in speed. There will be a marked trend toward articulation, separating the motive unit from the carrying.

On the other hand the forces moving against greater efficiency in road transport are considerable. The increase in the number of vehicles on the road may be more than seven million in the next twelve years. It will double the number now. If the motorways speed movement through the country, the congested approaches to the terminals in the towns may offset that advantage. The increase in utilization of the motive unit by articulation works for greater efficiency. Against efficiency works the increasing complexity and cost of the unit itself, the higher crew cost stemming from tighter organization of union membership and—be it said softly—wider observance of the law. In short, the efficiency of road transport will not surpass by much its present high level.

The Railway Revolution

The picture of the railways emerging from the age of the Stephensons is very different. Here the technical advances are revolutionary. The advances are under four principal heads: traction; design of rolling stock; marshalling yards; and terminals. There are indeed other things which follow such as stronger track, better signalling, automatic train control to suit the higher speeds but they are not fundamental to the end product which is a revolution in service.

Of the four heads the change from steam to electric or diesel traction is the most publicized and the least important in terms of service. The railways' problem is not so much to accelerate their already high speeds over the line but to reduce the time spent at a standstill at rail terminals and in marshalling yards. The principal advantages of the change of traction are economic. The capacity of the diesel and still more the electric, to work for days virtually without intermission, the greater reliability especially of the electric, the higher tractive effort will be translated into lower cost.

Rolling stock marches with the times in three ways. Firstly, to match the higher speeds; secondly, to overcome the disabilities of the road/rail transfer; thirdly, to eliminate the damage in shunting. Higher speeds mean heavier braking. The distance required to stop goes up with the square of the speed. The wagons must, therefore, share more in the stopping. A programme to equip all wagons with brakes which can be controlled from the locomotive is in course of being carried out.

Figure 4. Canadian National Railways mounted trailer being backed on to flat rail wagon for 'piggyback' transit

Road/rail transfer is the nub of the matter, the key to the railways continued existence. Economically they cannot continue to exist on the bulk traffics alone. They must secure a higher share of

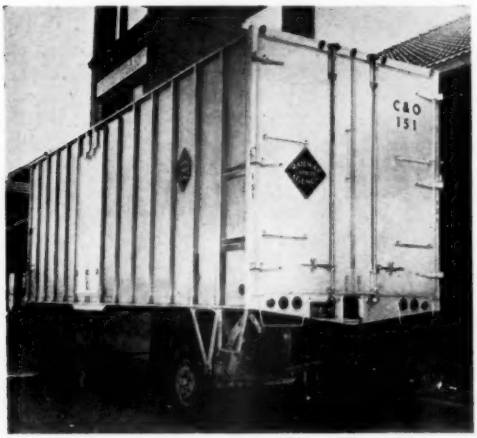


Figure 5. Prototype ' amphibian ' road-rail vehicle under development in this country (by courtesy of Chesapeake & Ohio Railway Company)

the consumer goods which require transport not siding to siding, but door to door. There are many devices to avoid transhipment at terminals under development at the moment, in two main classes; the container and the road/rail vehicle.

Container and Amphibian

The design of a container is partly a problem in design to suit individual trades: to tip for some bulk traffics, to discharge by compressed air for others (see for example *Figure 3*), to refrigerate; to carry greater weights. It is partly a problem how to transfer it between road and rail. Crane, mobile or overhead; fork lift truck; rollers and winches; lugs and ramps.

The road/rail vehicle may be either a ' piggy-back ' one (*Figure 4*) which climbs on to a railway chassis; or a true amphibian (*Figure 5*) which runs with the same chassis on road or rail. In America where everything including the railway's gauge is larger than life, the ' piggy-back ' is in wide use. Here with our gauge of 13 feet 6 inches, two feet less than the American's, the design of a ' piggy-back ' with a good pay load is difficult. The development of the amphibian is more likely. A prototype is under test. The design is being pressed and must be pressed with vigour. The upshot of these two things alone or in combination will be the ability of the railways to transfer full loads between road and rail in a negligible time at a negligible cost and with no risk of damage.

To safety in shunting, of which more will be said under marshalling yards, the contribution of the designer of rolling stock is the oleo-pneumatic buffer and automatic coupler (*Figure 6*), the effect of which will be simply—and no more need be said—to remove the risk of damage from shunting.

Marshalling Yards

The unit of the country's trade is less than a train load. The unit of a railway's operation is a train. Hence the device of marshalling yards for combining and dispersing the smaller units into and out of the larger—a necessary evil up to a point and that point is now several places away from where it was. Our forefathers could put no more than a thousand wagons a day into a marshalling yard of some seventeen sidings. That was the reasonable limit which an engine could do shunting on the flat. As traffic grew to the point where there were fifteen hundred, two, three, five thousand wagons a day passing through a junction, there was nothing our forefathers could do but add to the one yard at that junction, a sixth, a seventh and so on. At one such concentration a little less than five thousand wagons a day entered and left the area; we shunted over twelve thousand. Each wagon on average was shunted nearly two and a half times. Techniques have advanced to the point where we can shunt

Figure 6. Automatic coupling and oleo-pneumatic buffers which are under test on British Railways' freight wagons

over five thousand wagons a day in one yard. In the case quoted, 15 yards were combined in one main throughput yard and two storage yards.

The uses of these techniques are firstly to simplify the complex system of yards in the principal areas of production and consumption; and secondly to close or to reduce to their proper function of local distribution the yards intermediate on the line of route. In the first sense 15 yards in the East of London, 12 in Sheffield, become one in each case. In the second sense at Peterborough four yards were closed in 1953, three more are planned for closure. One region has closed 40 per cent of its yards since the war, another 30 per cent are under the hammer; 70 per cent in all. There is no better way of avoiding delay to wagons in a marshalling yard than by abolishing the yard which delays them.

Marshalling yards have a reputation for damaging wagons and their contents—exaggerated like many a canard but true up to a point. Apart from the oleo-pneumatic buffer being fitted to the wagons there are new techniques being developed in the yards themselves with railbrakes which control the speed of a wagon within very narrow limits. Each wagon determines by radar the degree of control which the railbrake must exert. In a few years damage in marshalling yards will be a thing of the past.

Terminals

Here there are three main streams of thought. Firstly, the new techniques of road/rail transfer by new types of container or by road/rail vehicle will accelerate and simplify the transit. Secondly, by virtue of these new vehicles the railways will no longer be tied to particular terminals. Any siding with a bit of tarmac alongside—for speed preferably alongside a marshalling yard—will do for a crane, fork lift truck or other device to transfer containers or for the road/rail vehicle to hop off or on to the railway.

Thirdly, for small consignments which must be sorted the concentration into terminals 20 or 30 miles apart will have the effects, firstly of making mechanization worthwhile with its benefits of speed and low costs; secondly of making train loads start from the terminal with benefits in speed, by avoiding marshalling yards and, thirdly of reducing damage in handling by the use of conveyors. In this sense one region has recently announced a plan for concentrating the work of 170 terminals in 48.

The end product in transit is something like this:

Time available (collection to delivery)	6 p.m. to 8 a.m.
Collection and transfer to rail, 1 h	7 p.m.
Marshalling and waiting departure, 1 h	8 p.m.
Journey up to 400 miles @ 50 m.p.h.	4 a.m.
Marshalling and placing, 1 h	5 a.m.
Spare 2 h	7 a.m.
Transfer to road and delivery	8 a.m.

The transformation of the railways in reliability, in overall speed, in safety, can hardly be exaggerated. If it is true that their present failures in these respects stem from the delays and risks in marshalling yards and terminals, then such a radical plan to avoid both in large part and to take the risk of delay or damage out of the remaining fraction must restore the railways to their old position of competing on equal terms of efficiency with any other form of transport. But—and it is a big but—there are two conditions precedent: courage to spend the very large sums of money on their freight business and the wit to solve the problem of the transfer between rail and road.

Finally—and this is the only pie in the sky—beyond the scope of this paper in time is the prospect of automation for the railways with their private tracks, their disciplined service, their high labour costs. To control loading, shunting and the movement of trains by push button is no idle dream.

The Upshot

The picture of inland transport after some 15 years have elapsed is of each form operating at a very high level of efficiency. Air, water, pipeline and transmission line will be doing important specialized work on the fringes of the national effort. The main stream will still be along the roads and the railways.

But unless there is effective action, there will be an enormous surplus of transport. National demand has increased by 45 per cent, C Licences up by 60 per cent, public road haulage vehicles by 50 per cent and trunk railways by 50 per cent.

The direction of action is clear. The end must be the preservation for the nation of the advantages of each form: to use the small flexible units of road and air for the small loads which need the flexibility of unregulated departures or arrivals or—in the air—exceptional speed and for all loads in areas from which the railways have withdrawn; to use pipelines for the bulk movement of oil from port to refinery; to use—and to use to near capacity—the railways which remain, for the main streams of medium and long distance traffic.

Left to itself, this will not happen. This is no time for a policy of *laissez-faire*. Unregulated competition in transport never has led to good service. Transport has to be governed and disciplined. Whether the need is for more regulation of competition or for more coordination or for integration is another matter. What is vital for the country is that the ' transport tap ' should not run to waste.

SURVEY

Lurgi Pressure-Gasification Process

The Lurgi process was originated in Germany before the second world war, and the plant at Westfield, on the borders of Fife and Kinross, whilst being the first in the United Kingdom, is in fact the sixth to be built in the world.

The great advantage of the Lurgi system is that low-grade non-caking coal can be used instead of high-grade coking coal. Instead of being carbonized, it is completely gasified by being treated with steam and oxygen under very high pressure, leaving only a residue of ash and no coke. The running costs of a plant such as this are low—some fifteen per cent less than in the more conventional plants—but this is partly offset by the very high installation costs; in this case the overall expenditure will be around £8,500,000.

The site at Westfield was chosen because of its proximity to the open-cast mines, which produce a low-grade non-caking coal with a twenty-five per cent ash content and a sixteen per cent moisture content. This coal is delivered directly to the plant from the mines on conveyor belts, and the ash is returned by conveyor belt to the land as a filler. (For a diagram illustrating the process see *Research* (1958 **XI,** p.157)

At the plant the coal is crushed and screened, and the larger coal fed to the Lurgi generators, whilst the smaller coal is used in the IGNA fluid stoker-fired boilers which raise the steam required in the Lurgi generators. Actually the Westfield gasification plant will not be using the whole output of the mines, and part of the coal is to be resold to the Electricity Board for generating electricity.

In the Lurgi gasifiers, which are large welded-steel cylinders, twenty-eight feet long and with an outside diameter of nine feet ten inches, the gas is made by reaction of the coal with oxygen and steam under a pressure of about twenty-five atmospheres. One gasifier will produce seven and a half million cubic feet of purified gas a day, using 240 tons of coal, 60 tons of oxygen and 220 tons of steam. The coal is totally reduced to gas and ash. Up to now two generators have been installed, and a third is following shortly. To begin with only two will be in operation at a time, and the third is a stand-by, so that when the plant goes into production at the end of 1960 or the beginning of 1961, fifteen million cubic feet of gas will be produced per day. More generators will be added later, and the plant will eventually produce thirty million cubic feet a day.

The oxygen required in the reaction will be produced on the site, by the Tonnox system in which air is compressed to liquid and the nitrogen boiled off by suitable reduction in pressure. There is storage capacity at Westfield for 550 tons of reserve liquid oxygen.

After leaving the generators, the gas will be cooled, and the crude benzole recovered. The gas will then be purified at the 'Benfield' plant, where two unwanted constituents, carbon dioxide and hydrogen sulphide, are washed out, and any remaining traces of hydrogen sulphide are removed in the high pressure tower purifiers shown in *Figure 1*. The purified gas has a heating value of 400 B.Th.U. per cubic foot, and this has to be boosted to 450 by the addition of butane. The gas will not be stored, as in most gas producing plants, but will be piped under pressure away from the plant to local grid

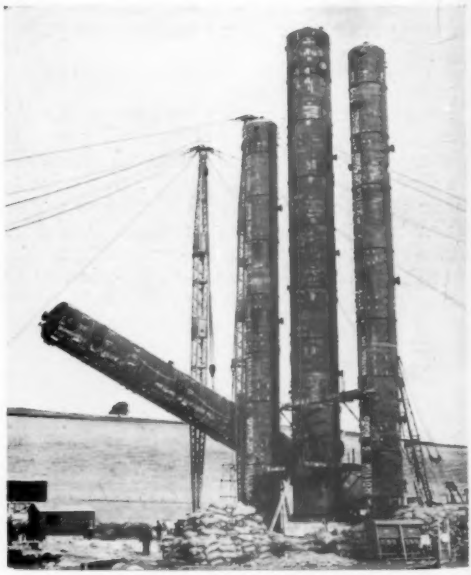


Figure 1. 'Bischoff' towers which will be used in the Benfield process for the purification of gas at Westfield

systems in Perth, Angus, Kinross, Fife, Stirlingshire, Lanarckshire and the Lothians. Steel pipes, twenty-five feet to forty-five feet long are to be used, these being laid along the top of the ground, welded on the spot, and then 'snaked' into four feet

deep trenches. The pipeline is designed to take a maximum working pressure 300 lbs/in².

Eventually one quarter of all Scotland's town gas will be produced at Westfield. This will involve the closing of a large number of small high-cost gas producing plants, and therefore additional economy of production.

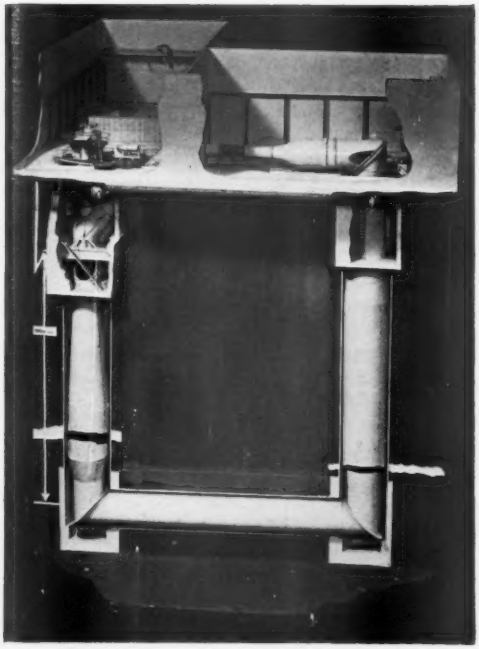


Figure 2. Model of 44 in. water tunnel
(Crown copyright reserved)

Ship Hydrodynamics Laboratory at Feltham

The ship hydrodynamics laboratory at Feltham, previously mentioned in *Survey**, was officially opened by H.R.H. the Duke of Edinburgh on October 19th. The main tank and the sea keeping basin are much as they were a year ago but the cavitation tunnel, which then was a large hole in the ground, is now completed and at work, (see *Figure 2*). The tunnel will test large propeller models up to 24 inches in diameter and also hydrofoil sections under controlled pressure conditions. The actual test section is 7 feet 4 inches long. The tunnel is designed for a maximum water velocity in the working section of 50 feet per second.

**Research, Lond.* (1958) 11, 488

One of the problems of operating water tunnels is to maintain a constant air content in the water. Dissolved air comes out of solution in the low-pressure high-velocity sections of the circuit and unless either removed or re-absorbed the water becomes full of bubbles and the propeller cannot be seen. Sea water, however, contains a considerable amount of dissolved air so that to simulate actual conditions it was decided to re-absorb the air during each circuit by taking the water deep down below the measuring section and subjecting it to high pressure while moving slowly. The lower horizontal limb of the tunnel is therefore 10 feet in diameter and 180 feet below the working section. The two vertical limbs are 87 feet apart from centre to centre.

One interesting disclosure was made at the recent press visit to the laboratory: it is now highly probable that a circulating water channel will be built. This would considerably increase the facilities of the laboratory as a circulating channel will enable small models of ships to be tested in moving water.

Productivity in European Railways

As a country or an area becomes more prosperous and its industries expand, it is necessary that the transport facilities expand correspondingly. The railway is the least adaptable form of transport, but the picture given to the European Productivity Agency, a branch of O.E.E.C.* shows that in the nine countries studied, Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Sweden and Switzerland, the railway networks increased their efficiency and their throughput during the years 1950 to 1957, though they appear, in fact, to have reduced the services offered.

The units used to express services *offered* were gross ton-kilometres carried or ton-kilometres offered (i.e. tons × distance travelled) for goods and seat-kilometres (seats × distance) for passengers. The units for services *sold* were passenger-kilometres, ton-kilometres or 'traffic units' which equated passenger-kilometres with ton-kilometres. The energy consumed by all types of engines was reduced to 'coal equivalents' and the production related to installed capacity, *i.e.* the total horse power at constant load of steam diesel and electric locomotives. In this way the figures from the different countries can be collated to give an overall picture.

It was found during the seven years under review the traffic had increased by 29 per cent. In spite of this the total number of man-hours worked per year has decreased by 9 per cent, due to technological progress, reorganization of networks, and schemes to

* Published in Productivity Measurement Review (1959) 18, 31

encourage increase of individual output (*e.g.* better industrial relations, bonus schemes, *etc*). Also, due to modernization, *i.e.* the change-over from steam traction to electric and diesel traction and the improved track equipment, the energy consumed during this period fell by 6 per cent, and the locomotive productivity index rose by 52 per cent. The number of seats or tons capacity offered has fluctuated during these years, but the productivity has increased by 20 per cent for passenger services and 25 per cent for goods, this being due to better planning and therefore less wasted capacity.

Sironized Wool

A new treatment for wool which combines a shrinkproofing and a fabric setting process to produce washable non-iron effects has been devised by C.S.I.R.O. Wool Research Laboratory, Geelong, Victoria, Australia. The wool is shrinkproofed by treating with potassium permanganate in a concentrated salt solution, followed by cleansing with sodium bisulphite with sulphuric or acetic acid, to remove the manganese dioxide formed on the wool. This treatment apparently prevents wool from shrinking due to felting during washing but does not prevent relaxation shrinking.

To achieve the non-iron effect the wool is first proofed against shrinking as above, brought up to pH 5, immersed in a cold one per cent solution of sodium bisulphite with a wetting agent and steamed damp on a blowing machine. Apparently it is important that the fabric should be at exactly the right pH as stiffening occurs if the pH is too high, even at pH 6, and if the pH is lower than 4, setting is not effective. This part of the treatment is intended to ensure that there is no surface distortion or puckering of the woolen or worsted fabrics after washing.

Recent Biological Symposium

Biological problems arising from the control of pests and disease was the subject of a symposium in London arranged by the Institute of Biology on October 1st and 2nd*. Such problems were broadly categorized as those arising in the course of treatment and those resulting from control. The former problems included the danger of insecticides to operators and consumers; however, suitable precautions resulted in little risk. By a process of accelerated evolution, our 100 harmful species of insects were not immune to certain synthetic insecticides. With bacteria of public health significance, the problem in hospitals was confined to some species which had become drug tolerant, drug-destroying or drug-dependent.

The control of specific diseases or pests permitted the evaluation of others; but the consequential effects of such control in tropical or underdeveloped areas could lead to over-population, malnutrition, soil erosion or the spread of a virulent epidemic of a reintroduced disease. Therefore a policy of education, rehabilitation and economic and agricultural development was a necessary adjunct. Certain treatment of plants and livestocks intended for breeding was bad policy if inferior characteristics were to be perpetuated. Breeding disease-resistant plants and livestock as an alternative means of control was said to be a slow and sometimes impossible process.

* The full Proceedings will not be published until 1960

CERN Proton Synchrotron

During the CERN Accelerator Conference (September 14–19) the Organization's twenty five thousand million electron volt proton synchrotron beamed its particles once round its 628 metres circumference.

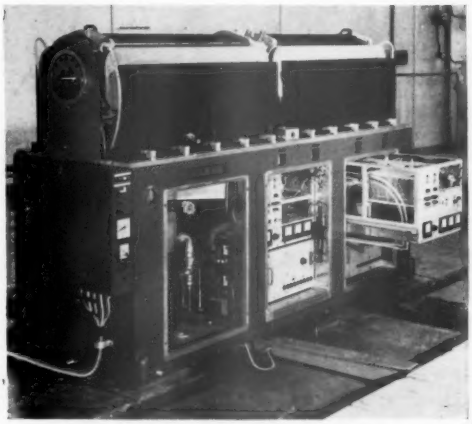


Figure 3. One of the 16 radio frequency accelerating units responsible for the gradual build-up of proton energy to 25 GeV (BeV) during the acceleration cycle of one second. In passing through these 16 units, i.e. per single turn around the 200 m diameter ring building, the protons gain an average of 50 to 54 keV. (Reproduced by courtesy of CERN)

The full acceleration cycle should be reached next year, and this will require one second, during which the protons will make 50,000 turns to produce 26 GeV proton energy. (See *Figure 3*).

SIR HENRY TIZARD,
G.C.B., F.R.S.

THE SCIENTIFIC ADVISORY BOARD of *Research* has suffered much in recent years through the deaths of founder members. We have lost our first chairman, Viscount Waverley, Sir Wallace Akers, Sir Alexander Fleming, and in the last issue we had to report the death of Sir Alfred Egerton. Now there is the dismal news that Sir Henry Tizard died on October 9th.

Tizard's father was in the Navy and became in due course Hydrographer to the Admiralty. The initial intention that Sir Henry should enter the Navy himself was frustrated by his eyesight, and instead he became a distinguished scientist–administrator for which his background and experience were peculiarly fitted. He was trained initially as a mathematician and a chemist, at Oxford with Sidgwick, and in Berlin under Nernst but it was in government work in the first world war that a peculiar and remarkable stamp was given to his career. In this the influence of Bertram Hopkinson was definitive.

Though Tizard occupied many offices, he was Secretary of the Department of Scientific and Industrial Research from 1927 to 1929, Rector of the Imperial College of Science and Technology from 1929 to 1942 and President of Magdalen College, Oxford from 1942 to 1946, it was as a co-ordinator and sharpener of Government scientific policy that he will be chiefly remembered. In the reference lists the important committees appear during and after the war; he was for instance Chairman of the Advisory Council on Scientific Policy and of the Defence Research Policy Committee from 1946 to 1952 and as such had great influence on the post-war national policy. Yet in a very real sense Tizard's most important work was done before the war. His strategic grasp and executive insight were of the highest importance to the country in these years and the harvest of his outstanding talents was really reaped during the war.

Tizard had a restless, quick, probing mind which was most stimulating. He had a downright and at times blunt form of expression which could be deceptive. His broad experience of the diffuseness of committees led him to express opinions in terms of clear specific propositions. But this by no means meant that he had a closed or prejudiced mind. If a contrary opinion was supported by clear evidence, he was the first to appreciate its value and after review of the opposing arguments, give it categoric support. He was a great clarifier of opinion. The Scientific Advisory Board will miss his grasp and his pungency.

BOOK REVIEWS

Aromatic Substitution: Nitration and Halogenation

P. B. D. DE LA MARE and J. H. RIDD

(vii + 252 pp; 10 in. by 6¼ in.)

London: Butterworths Scientific Publications; New York: Academic Press. $9·00, 50*s*

THIS is a most valuable monograph. The necessary amount of qualitative organic theory, including that of the transition state, is dealt with first and is quite adequate. Then follows an excellent account of the investigation of reaction mechanism, in which the criteria enabling one to decide on one of several possible reaction schemes are discussed. There is a short preparative chapter introducing fifty pages or so devoted to nitration studies of benzene compounds; the role of nitrous acid in those reactions in which it is involved is dealt with. Halogenation studies in the benzene nucleus comes next, and are introduced with a short preparative chapter. For the more complex compounds—diphenyls, polycyclic aromatic compounds, tropolones, *cyclo*pentadienes, and the principal heterocyclic substances, nitration and halogenation are considered together. There is an all too brief chapter on electrophilic displacement of groups other than hydrogen—the authors have limited their choice of examples to those where the reaction path is reasonably certain. Finally, there is an account of the results of recent molecular orbital and free energy calculations. At the present time it would generally be admitted that the agreement with experiment is only qualitative.

The account of nitration is in many ways the most satisfying, but only because the picture is so much more complete than with halogenation. Unfortunately the text has no author index, despite a wealth of references which are right up to date. This is a stimulating book, and it is to be hoped, as hinted at in the introduction, that the authors will follow it up with another dealing with analogous reactions. J. S. WHITEHURST

A New Method in the Theory of Superconductivity

N. N. BOGOLIUBOV, V. V. TOLMACHEV and D. V. SHIRKOV

(iii + 120 pp; 9¼ in. by 6 in.)

New York: Consultants Bureau; London: Chapman & Hall. $5·75, 50*s*

RECENTLY there has been much activity towards the development of a basic theory of superconductivity. The problem is this: given a set of equations governing the movements of electrons in an idealized metal, can one deduce from them the peculiar phenomena associated with perfect conduction near absolute zero temperature? Mathematically this problem is not an isolated one, but falls into a whole class of problems ranging from the theory of elementary particles to the properties of stellar atmospheres, which concern the theory of systems with very many degrees of freedom. Such a theory has the somewhat awkward feature that at the present time no one has been able to establish any rigorous mathematical method of proving whether a given approximate solution does in fact follow from the equations of the theory. Faced with this uncertain situation the physicist is forced to rely on his intuition and on experimental tests to support his approximations.

It is perhaps because of this vagueness that a peculiar air, almost of dogma, seems to have grown up around recent theoretical work, notably in the theory of superconductivity; and this would appear to have affected the publishers of the above volume, who describe it on their rather pleasing, if colourful, dust jacket by the phrase 'in this unprecedented *complete* solution to the perplexing problem of constructing a microscopic theory of superconductivity'. In fact the book is a translation of recent research work, much of it previously unpublished, which appeared originally in book form (in Russian) in 1958. Two introductory chapters summarize previous work by N. N. BOGOLIUBOV in this field, and the remaining six chapters of the book are devoted to an account of some highly technical development of this work. The work is still in progress, as may perhaps be judged by the appearance of subsequent modifications of the theory which have been introduced by Bogoliubov to deal with some problems of the electromagnetic properties of superconductors.

In view of this it is rather a pity that the publishers did not note the recent publication of an English translation of this same work which has already appeared, in slightly shortened form, in the *Fortschritte der Physik*. They could, perhaps, have avoided this duplication in a field which is still developing rapidly. S. DONIACH

Hyperstatic Structures: Volume I

J. A. L. MATHESON

(xv + 474 pp; 10 in. by 6 in.)

London: Butterworths Scientific Publications. 90*s*

THE aim of this introduction to the theory of statically indeterminate structures is to treat the methods of analysis and the several theorems on which these methods are based as an organic and coherent whole. It describes both classical theory and more modern methods and demonstrates how the various techniques are interrelated.

The first chapter deals with the diverse types of redundancy, the second with energy theorems, the third and fourth with general theorems applicable to linear structures, the fifth with moving loads, the sixth with rigid joints, the seventh with arches, the eighth (by Dr N. W. MURRAY) with stability and the ninth (by Dr R. K. LIVESLEY) with matrix methods. The analysis relates chiefly to perfectly elastic structures, but there are passing references to limit design and a short descriptive account of collapse loads of structures associated with the development of plastic hinges. The emphasis throughout is on practical design problems, and each section is so

lavishly illustrated by worked numerical examples that it is a little surprising that still more examples are to follow in a second volume.

The book should certainly be of value in structural design offices in that it provides a wide range of examples of the several methods of analysis now in use. The interrelation of these methods could however have been better illustrated by application of different techniques to one or two more carefully chosen examples.

The work is well printed and bound and it appears free from misprints. The title is regrettable; the adjective 'redundant' and the phrase 'statically indeterminate' are well known, and there is no need for a neologism which has yet to gain a place in the dictionary. When the hyperclassical lexicographer does admit 'hyperstatic', can he define it otherwise than 'extremely stationary'?

H. L. Cox

The Physics of Electricity and Magnetism

W. T. Scott

(*xvi+635 pp; 9¼ in. by 6 in.*)

New York: John Wiley; London: Chapman & Hall. $8·75, 70*s*

The chapter headings of the book are: charge, field and potential; metallic conductors; dielectric materials; further topics in electrostatics; steady currents; phenomena related to electron levels; the magnetic field; magnetic materials; alternating currents; electromagnetic radiation.

The text, written with great pains by the author to provide 'a sound, logical basis for understanding the physics of electricity and magnetism' seems to fail on three grounds. Firstly, the author attempts to include in one volume three subjects—electromagnetism, circuit theory and the properties of materials—which because of the wide differences in the mathematical processes involved are best treated quite separately. Secondly, his treatment of electromagnetism follows the old conventional lines, which it is believed experience has shown to be unsatisfactory. Thirdly, the first essential of any textbook is that it should be interesting, whilst here a conscientious effort on the author's part to anticipate in the early chapters all a student's possible questions has merely made the text long and tedious.

The physical laws of circuit theory are few and easily understood and the mathematics involved is relatively straightforward. The treatment given by the author, though sound is elementary, but the fifth and ninth chapters devoted to this subject by being divorced from the rest of the text effectively constitute two long and undesirable digressions from the main theme. His atomic approach to the electric and magnetic properties of materials is largely descriptive and free from mathematical content. Here again the treatment is elementary but now the material is scattered about the text. Atomic physics is not *necessary* in macroscopic electromagnetism. In the reviewer's opinion its effect is to confuse rather than clarify. For this book is primarily one on electromagnetism, and to develop this difficult subject as regards relativity, clarity and continuity are each essential.

The first reason given in the preface for the appearance of the book is the author's 'dissatisfaction with available texts in the field'. Bewilderment on the part of the student and dissatisfaction on the part of the teacher are commonplace in electromagnetism, and it is quite remarkable how many authors have been inspired by an initial dislike of conventional texts to produce a text of their own which is essentially the same as those they originally disliked. One of the earliest and best conventional treatments is given in Jeans' classic *Electricity and Magnetism*. Of the numerous books which have appeared since Jeans', only two have offered a radically new approach. Both were written by professors of physics and both introduce *naturally* rationalized M.K.S. units. One begins by postulating Maxwell's equations on the first page. This leads to a stimulating approach, but one unsuited to most beginners. The other, *Physical Principles of Electricity and Magnetism* by Pohl published in 1930 begins with current electricity and proceeds to electrostatics via the parallel-plate capacitor and to magnetism via the long solenoid. When one is familiar with Pohl's approach, the whole subject becomes so clear and convincing that one can never again view the conventional pithball treatment without exasperation. The reviewer believes that the universal difficulties of electromagnetism are largely inherent in the conventional treatment and that they can never be eliminated by any rewriting of this treatment, however painstaking.

As an excellent illustration of these difficulties, the first three pages of the present text may be quoted. Here a beginning is made with Coulomb's Law in rationalized M.K.S. form

$$F = \frac{q_1 q_2}{4\pi \epsilon_0 r^2} \cdot I_r$$

followed by a discussion of units and dimensions, which must be quite incomprehensible to the beginner. From this inauspicious start recovery may be well nigh impossible.

G. V. Jones

Techniques of Polymer Characterization

P. W. Allen (*Ed.*)

(*xvi+256 pp; 8¾ in. by 5½ in.*)

London: Butterworths Scientific Publications; New York: Academic. 50*s*, $9·50

The ever-increasing complexity of high polymer science makes any competent review of the subject welcome. In this book, a group of specialists have described the methods by which some insight into the physical and chemical nature of given polymers may be obtained. After an introductory chapter by the editor on definitions, and a brief discussion of polymer solution, the fractionation of high polymers is considered by R. W. Hall. The following four chapters describe molecular weight determination, including osmometry by H. T. Hookway, light-scattering techniques by F. W. Peaker, viscometry by P. F. Onyon, and a short but interesting chapter by D. F. Rushman on other methods, including ebulliometry and cryoscopy. Techniques of end-group analysis are covered by G. F. Price, and, finally, the characterization of block and graft copolymers is considered by R. J. Ceresa.

In each case the pattern of the chapter is similar. A more or less brief discussion of the relevant theory, in some cases rather over long, is followed by an account of the techniques involved. Here, no complete description of available methods has, in most cases, been attempted; instead each author has tended to describe his own experience. Thus the book is rather unbalanced as a critical account of techniques, but at the same time many useful points of great practical value are noted. Of special interest are the sections on the interpretation of results, particularly those discussing molecular weight determinations, both absolute and relative to each other. For anyone new to high polymer work, the book can be read with great value, bearing in mind the occasional lack of perspective. Polymer scientists, interested in either academic or applied aspects of this fascinating subject, will find much stimulation, and, sometimes, controversy, amongst its pages. C. A. FINCH

The Transuranium Elements

G. T. SEABORG

(xx+328 pp; 9½ in. by 6 in.)

London: Methuen. 50*s*

PROFESSOR GLEN T. SEABORG's 1957 Silliman Lectures at Yale University had for subjects ' The Plutonium Story ', ' Chemical Properties of the Actinide Elements ', ' Nuclear Properties of the Transuranic Elements ', and ' Future Synthetic Elements '. In this book they appear as four chapters, but for this purpose the first three lectures have been so much enlarged as to bear only a formal relationship to the originals. The author has used some of the expansion, especially of the first subject, to mention most of the names of those who contributed to the discoveries. He is seen to be discharging what he has accepted as a duty. At this it is difficult to cavil, but the strings of names which have been brought into the text, and the lists of others which occupy whole pages, do handicap the style and presentation. Nevertheless the powers of the author and the coherence of the subject are such that the plutonium achievement and its significance emerge clearly.

The lectures were not addressed to specialists and in these chapters the approach remains broad and general, though a good deal of detail has been introduced and parts of Chapter 3 are not easy. But the clarity and verve of the writing carries the reader through or round the difficulties according to his capacity or diligence. There is a clear picture of the chemistry of the actinides, that new rare earth series in which the 5*f* electron shell is filled, appropriately illustrated with a colour plate of the ions in their various oxidation states.

Up to this point the going has been easy, but the account of nuclear properties of the transuranium elements which deals with nuclear reactions, thermodynamics, the unified nuclear model, decay energies and rates, even–even and odd–even nuclei, and induced and spontaneous fission is more difficult. However, for those seeking an understanding of current views of the nucleus, it is difficult to think of a more rewarding 98 pages anywhere.

The last chapter is an attractive piece of writing. Little expanded, one imagines, from the text of the lecture, it looks into the future of element synthesis through eyes of great experience. Prediction, production and detection are discussed separately and knitted together by an exciting account of the synthesis of mendelevium. Those who have heard the author lecture will want to read this book. It is recommended to the non-specialist and could be read with great profit by all students of physics and chemistry.

The production is exemplary and the illustrations most helpful. P. L. ROBINSON

The Cathode-Ray Tube and its Applications *(3rd Edn)*

G. PARR AND O. H. DAVIE

(xii+433 pp; 8¾ in. by 5½ in.)

London: Chapman & Hall. 50*s*

THE first edition of this book appeared in 1937, but enormous advances have been made since then in the apparatus and techniques of oscillography. This third edition is therefore virtually a new work. The authors' aim has been to provide a handbook for anyone who has to use a cathode ray tube. The main emphasis is on the basic principles of the various circuits used with the cathode ray tube and not on the practical details. Consequently a reader expecting to find a discussion on commercial oscilloscopes will be disappointed.

Following a short but interesting historical introduction, a single chapter is devoted to the cathode ray tube itself. The authors have wisely left detailed electron optics to more specialized works and have included here only a simple description of the construction and properties of modern cathode ray tubes.

More than half of the remainder of the book describes the principles of the circuits essential for the exploitation of these properties. The chapter on amplifiers is an excellent introduction to the principles of electronic amplification and is the best section of the whole book. The chapter on time bases contains a clear description of many different circuits but would probably have been more helpful had the approach to the practical merits and disadvantages of these circuits been more critical. The ' Miller effect ' is mentioned several times in this section, but is never clearly described or defined.

Many of the applications which have been found for this versatile instrument are described in the last part of the book, with the emphasis remaining on principles rather than on practical detail. With an average of about two and a half pages for each topic, the authors can only indicate the most important features of each application, but for those readers who require more information the chapters each conclude with a full list of references.

The book contains a large number of illustrations, most of which are circuit diagrams. These are particularly clear and easy to follow. There seems little justification for the inclusion of some of the photographs, however. The general presentation is excellent, with the text well printed on good paper, strongly and attractively bound to produce a useful and interesting volume. H.N.D.

Applied Hydrodynamics

H. R. VALENTINE

(viii+272 pp; 8¾ in. by 5½ in.)

London: Butterworths Scientific Publications. 50*s*

THIS is a textbook intended for the student just beginning his study of ideal fluid flow; the subject matter, concerned in the main with two dimensional problems, is followed by the usual route up to the stage of quite complicated examples of conformal transformation. So much might be said of many books which have from time to time been published, and no doubt may be said of many more yet to come. For the subject is notoriously difficult to 'put across', involving unusual concepts and a host of dubious assumptions, and the new approach is constantly sought.

Most textbooks take to the main road of mathematics, laid down these many years, cutting remorselessly through the tangled bush, bridging pitfalls, and so on. The speed of passage, and the turnings taken, vary according to the driver, and not all pursue the chase until the sign is seen that there are 'road works ahead', but the road provides the surest passage to this end. Many there at work have come this way, and know little of the countryside about them.

Other texts which seek to give the physical background of the study might indeed be said to leave the roads, and to explain the geography of the vast terrain they penetrate. But there are a precious few indeed which venture far afield, and this book although it sets out here and there, quickly returns to the tar-macadam. For instance, at the junction where the roads of irrotational and rotational flow part company, no persuasive reason appears why we should choose the former: we merely find ourselves upon it.

The chief merit, and indeed the chief theme, of this book is an unusual one and for this reason it is much to be commended. It could be called a 'driving manual', showing the young traveller how to make headway along the road of mathematics, which itself is not all that smooth: it succeeds in this by full and well-ordered argument, and a host of examples. These include many, for example, on flows with free-streamlines, which are usually beyond the scope of books such as this. The careful attention to practical methods of streamline construction, and in particular to relaxation methods, is also a novel and valuable feature.

The material chosen is particularly relevant to the study of hydraulics; but the aeronautical engineer is also a prospective client, though maybe his attention these days is mostly diverted away from 'classical' hydrodynamics. The student of mathematics, too, should find much here of value unless his false pride keeps him with the terser texts. However, to all alike, it can be wholeheartedly recommended. The book is beautifully produced, and the illustrations not only clearly drawn and printed, but well designed. T.N.

LETTER TO THE EDITOR

Theory of Friction of Polymers

SIR,

Theoretical approaches to the friction of polymers have usually been in terms of the adhesion mechanism originally developed for metals[1]. Contact is made at the tips of asperities and the frictional force is proportional to the real area of contact between asperities. For polymers, whose deformation is visco-elastic, it has been found that the number of such asperity contacts is important[2–4]. However, only two simple cases have previously been considered, corresponding on the one hand to contact between single asperities[5] and on the other to contact between separate asperities whose number remains constant or increases with increasing load[3,4].

In the following note a more comprehensive approach is suggested in which the two cases mentioned above appear as the two extremes of a more general type of behaviour. The frictional model involving multiple asperity contact is thought to be adequate provided flattening of asperities under the influence of the load is taken into account. If this occurs to an appreciable extent the distinction between single and multiple asperity contact begins to disappear and this is also true if the overall contact area is very small as, for example, with a pair of crossed fibres of small diameter. In this treatment the multiple asperity theory is modified to include the intermediate region between multiple asperity contact as normally understood and complete contact such as occurs between single asperities. This transition region appears to be of some importance in the friction of polymers, as is shown, for example, by an analysis

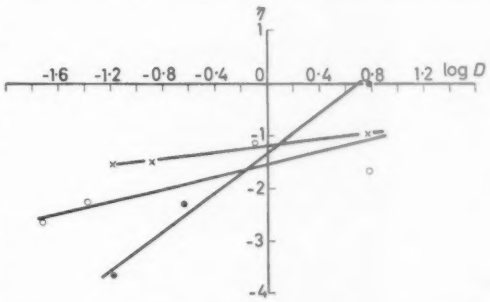


Figure 1. Undrawn nylon 66 (○), *crossed fibres with D 0·019 mm, 0·042 mm and 0·8 mm and sphere on flat D=6·0 mm; undrawn polythene* (×) *surfaces D=0·065 mm, 0·13 mm and sphere on flat D=6·0 mm; undrawn P.V.D.C.* (●) *surfaces, crossed fibres D=0·065 mm, 0·228 mm and sphere on flat D=6·0 mm*

of the results of M. W. Pascoe and D. Tabor[5]. They used crossed single fibres of different diameters or a sphere sliding against a flat. With nylon the diameters ranged from 0·019 to 6 mm and the loads from 10^{-6} to 10^{4} g.

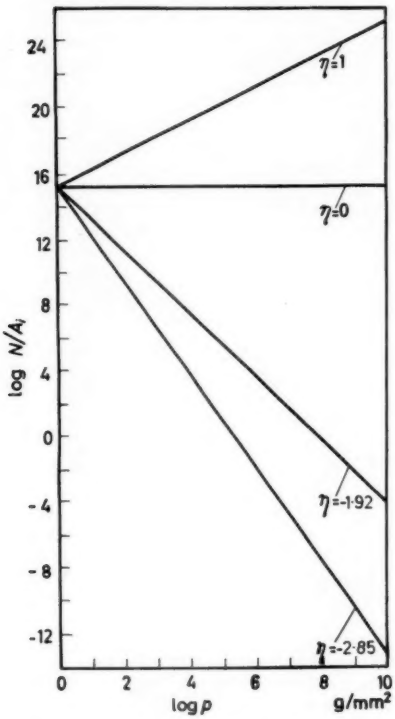


Figure 2. Clean undrawn nylon 66 surface: N/A_i versus p

Suppose the total number of asperity contacts is represented by

$$N = C A_i p^{\eta} \qquad \ldots .(1)$$

where A_i is the indentation area between the crossed fibres—or the sphere and flat—and p is the average pressure over the indentation area. C and η are constants. The expression previously derived[2,3] for the coefficient of friction is

$$\mu = f \alpha (N/W)^{1-m} \qquad \ldots .(2)$$

where W is the total load and the frictional equation for a single asperity contact with load w is

$$F = \alpha w^{m}$$

where $f = \overline{w^{m}}/(\overline{w})^{m}$

and is nearly unity. Substitution of (*1*) in (*2*) gives

$$\mu = f \alpha C^{1-m} p^{(\eta-1)(1-m)} \qquad \ldots .(3)$$

where[5] $p = W/A_i$

$$= W/K W^{m} D^{2(1-m)} \qquad \ldots .(4)$$

D being the diameter of the fibre or sphere and it being assumed that A_i is proportional to W^{m}, *i.e.*, m defines the deformation characteristic of a single asperity and also of the bulk polymer, the latter having been obtained experimentally by P and T for each of the polymers considered.

From (*4*)

$$p = \frac{1}{K}\left(\frac{W}{D^2}\right)^{1-m} \qquad \ldots .(5)$$

Thus from (*3*) and (*5*), μ is a monotonic function of W/D^2. The coefficients of friction obtained by Pascoe and Tabor for nylon, polythene, polyvinylidenechloride (P.V.D.C.) and polytetrafluoroethylene (P.T.F.E.) over an enormous range of loads and diameters were all represented as monotonic functions of W/D^2, this giving the best correlation of the results. The above argument suggests a simple basis for the correlation.

Substitution of (*5*) in (*3*) gives

$$\mu = \frac{f \alpha C^{1-m}}{K^{(\eta-1)(1-m)}}\left(\frac{W}{D^2}\right)^{(\eta-1)(1-m)^2} \qquad \ldots .(6)$$

If η is constant this equation is identical in form with that used by Pascoe and Tabor. For all their results their equations imply:

$$(\eta-1)(1-m)^2 = -(1-m) \qquad \ldots .(7)$$

For nylon m is 0·74 and η is −2·85. Negative values of η are also obtained with the other polymers when the results are treated as above[5]. These values are given later.

A negative value of η implies that N/A_i (*Equation 1*) falls with pressure. It would be expected that η would be a minimum for $N=1$, since for multiple asperity contact increase of A_i would increase N, whereas A_i increases without any corresponding increase of N for single asperity contact. Expressing N as a function of pressure alone from (*1*) and (*4*)

$$N = C K^{1/(1-m)} D^{2} p^{\eta+[m/(1-m)]} \qquad \ldots .(8)$$

For $N = 1$, $\qquad \eta = m/(m-1) \qquad \ldots .(9)$

Since N cannot fall below unity with increase of pressure *Equation 9* defines a minimum possible η value. This minimum η implies complete contact between the surfaces, *i.e.*, contact as complete as that between individual asperities. At this value of η, N is only unity for a particular value of D—assuming C and K are constant—and this D value must correspond to equal curvature of asperity and fibre. If the curvature of the asperities exceeds that of the fibre the N value corresponding to a minimum η is constant but greater than unity.

Equations 7 and *9* are identical, so that complete contact was assumed in obtaining the equations used to correlate the results[5]. For incomplete multiple asperity contact it would be expected that η would tend to become less negative than the minimum possible value, and *Equation 8* shows that in this case N increases with p. It is of interest to see how η varies as the diameter is increased with crossed fibres or a sphere against a flat. This variation may be obtained from Pascoe and Tabor's experimental results and *Equation 6* by plotting $\log \mu$ against $\log (W/D^2)$, since

$$\log \mu = \log\left(\frac{f \alpha C^{1-m}}{K^{(\eta-1)(1-m)}}\right) + (\eta-1)(1-m)^2 \log\left(\frac{W}{D^2}\right) \qquad \ldots .(10)$$

The slope of the lines gives η, m being known, so that η may be plotted against log D as in *Figure 1*. For nylon and P.V.D.C. values of m of 0·74 and 0·83 equal to those obtained in the deformation experiments[5] were used. For polythene the corresponding m is 0·74, but this gives η values substantially lower than the minimum possible indicated by *Equation 9*. However assuming $m = 0{\cdot}6$ for polythene, the value of η for the 0·065 mm diameter crossed fibres becomes equal to the minimum value of $m/(m-1) = -1{\cdot}5$, suggesting approximately single asperity contact. The variation of η with log D is therefore shown for polythene in *Figure 1* with $m = 0{\cdot}6$ instead of 0·74, the difference suggesting a greater degree of elastic deformation in the friction experiments than in the deformation experiments of Pascoe and Tabor.

In all cases in *Figure 1*, η is negative and it becomes less negative with increase in diameter, indicating a transition from complete—possibly single asperity—contact to incomplete multiple asperity contact. Multiple asperity contact, as normally understood, only occurs when the number of contacts remain constant or increase with increasing pressure ($\eta \geqslant 0$). With P.V.D.C. this condition is just realized for the 6 mm diameter sphere sliding against a flat ($\eta = 0$).

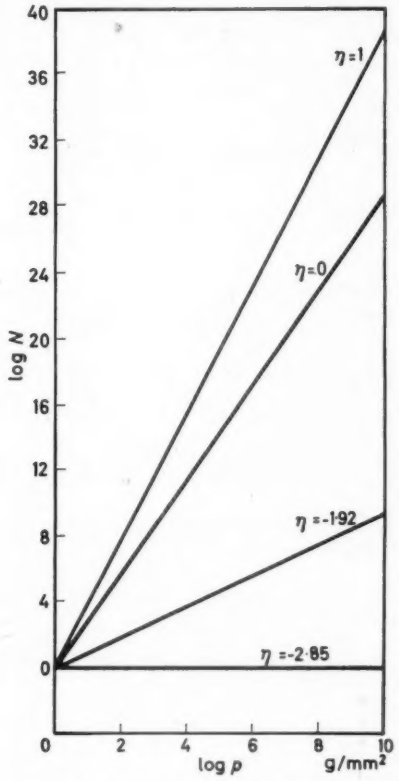


Figure 3. Clean undrawn nylon 66 surface: N versus p

In *Table 1* are shown some of the more salient features of the analysis of Pascoe and Tabor's results. Their interpretation of these results is seen to imply complete

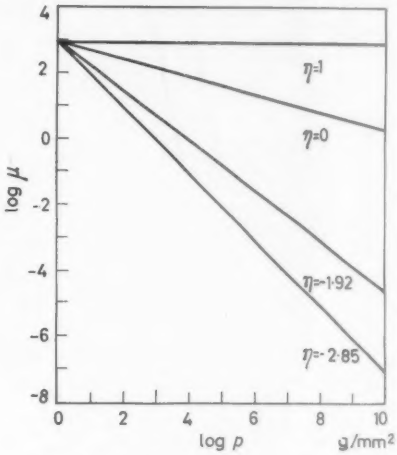


Figure 4. Clean undrawn nylon 66 surface: μ versus p

contact—η a minimum—in all cases, this giving a crude fit with the experimental results. A rather better fit is obtained by averaging the η values shown in *Figure 1* corresponding to various diameters for a given polymer.

Table 1. Analysis of Pascoe and Tabor's results

Polymer	m	η*	η†	η‡
nylon	0·74	−2·85	−2·85	−1·92
polythene	0·6	−1·5	—	−1·32
polythene	0·74	−2·85	−2·85	—
P.V.D.C.	0·83	−4·88	−4·88	−1·99

*Minimum from *Equation 9* †Implied by Pascoe and Tabor
‡ Average of those in *Figure 1*

These average figures are given in the final column of *Table 1* and they imply that contact in general is neither complete, as with a single asperity, nor as incomplete as in the usual multiple asperity picture, but intermediate between the two.

The variation of N/A_i, N and μ with p for values of η varying from the minimum possible to unity is shown in *Figures 2*, *3* and *4*. These figures were obtained by calculation from *Equations 1*, *8* and *3* using constants applicable to nylon crossed fibres of small D, ($f\alpha = 0{\cdot}1$, $C = 1{\cdot}66 \times 10^{15}$, $K = 4{\cdot}95 \times 10^{-4}$).

Diagrams such as these cover all the possibilities that are likely to arise in practice in relation to the friction of polymers. For example it is of interest to consider Thirion's law for the friction of rubber in relation to the present treatment. Thirion's law states

$$1/\mu = a + bp \qquad \ldots .(11)$$

At low pressures this equation shows μ is constant, *Equation 3* giving this result when η is 1. At high pressures *Equation 11* gives $\mu = 1/bp$, *Equation 3* giving this result when

$$(\eta-1)(1-m) = -1$$
$$\eta = m/(m-1) = -2$$

if m is 0·67. This equation is the same as *Equation 9* defining the minimum possible η value. (It should be noted that if a flat rubber surface is used sliding against another flat surface $A_i = A_a$ where A_a is the apparent contact area which may be fixed. This does not alter *Equation 3*

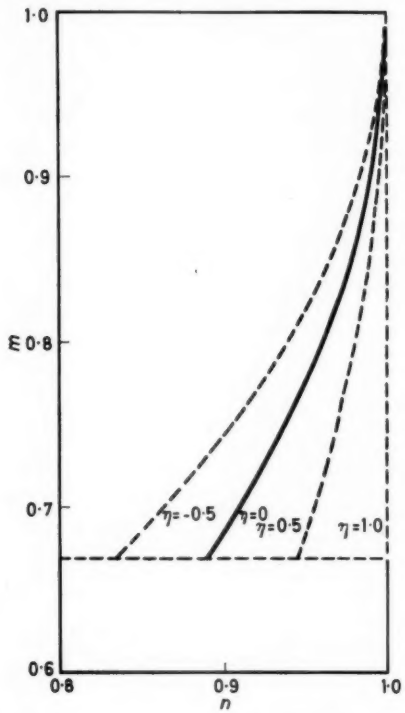


Figure 5. m versus n for different values of η

p now being equal to W/A_a). Thus at high pressures η falls to the value it would have for a single asperity, which implies complete contact between the surfaces. Thirion's law thus appears to cover the entire range of η values, η initially falling from one to zero (multiple asperity contact) as the load increases. At higher loads multiple asperity contact does not rigourously apply, the transition region occurring whilst η falls from 0 to -2. At η equals -2 contact between the surfaces is as complete as it is between single asperities.

It would seem that the frictional behaviour of polymeric materials such as nylon, polythene and P.V.D.C. is to some extent rubber-like in so far as negative values of η are common, *i.e.*, the pressure may often be sufficient to cause nearly complete flattening of asperities as with rubber at reasonably high pressures[6].

Equation 6 may be expressed in the well known form

$$F = AW^n \quad \ldots .(12)$$

where

$$n = 1 + (\eta - 1)(1 - m)^2 \quad \ldots .(13)$$

Consider n values approaching unity (Amonton's law); *e.g.* for $n = 0{\cdot}99$, *13* implies that for $\eta \geqslant 0$, $m \leqslant 0{\cdot}9$ *i.e.*, for the usual multiple asperity picture to apply m must be 0·9 or less. In view of the inherent variability in the frictional measurements between metal surfaces it would be impossible in general to distinguish between n values of 0·99 and 1, and this suggests that for metals the assumption of purely plastic deformation may not be always justified. The very critical way in which m drops below unity for slight decreases of n from unity is illustrated in *Figure 5*—values being calculated from *Equation 13*—which shows that care must be used in deducing the mode of deformation of asperities from experimental results giving n values in the region of unity. In the special case where η is 1, n equals 1 for all practical values of m, this being the case considered by J. F. ARCHARD[4].

η Appears to be a deformation index determining the real area of contact between surfaces. It is a function of the change in the distribution of asperity contacts with pressure over the region of contact. The value of η depends on whether the number of asperity contacts per unit indentation area increase (η positive) remain constant (η zero) or decrease (η negative) with increasing pressure. Maximum and minimum η values occur due to the existence of two opposite limits in the geometrical conditions between the surfaces within the indentation area.

The author wishes to thank Mr J. G. Hutton for computational assistance.

Yours faithfully,

J. D. HUFFINGTON

British Jute Trade Research Association
Dundee

30th June 1959

References

[1] BOWDEN, F. P. and TABOR, D. *Friction and Lubrication of Solids* Oxford: Clarendon Press, 1954
[2] HUFFINGTON, J. D. *Research, Lond.* (1957) **10,** 163
[3] — and STOUT, H. P. *ibid.* (1958) **11,** 450
[4] ARCHARD, J. F. *Proc. Roy. Soc. A* (1957) **243,** 190
[5] PASCOE, M. W. and TABOR, D. *ibid.* (1956) **235,** 210
[6] DENNY, D. F. *Proc. phys. Soc. B* (1953) **66,** 721

NOTICE TO CONTRIBUTORS: All communications intended for the Editorial Department should be addressed to The Editor, *Research*, 4 and 5 Bell Yard, London, W.C.2. Contributions sent on approval will be carefully considered by the Editor; but no responsibility whatsoever can be accepted for manuscripts, although every effort will be made to return them if a stamped addressed envelope is enclosed for that purpose. The copyright of all contributions for which a fee is paid shall belong to Butterworths Publications Ltd, together with the right of republication. The Editor reserves the usual right to make alterations in the text of articles accepted for publication. The author will receive galley proofs and if alterations have been made every effort will be made to obtain the author's prior agreement. Unless special arrangements are made, contributions are only received and considered on these conditions.

as

ric
to
of
ent
ith

rm
12)
13)
.g.
the
or
nal
be
ues
the
be
ops
is
om
in
om
of
all
by

the
of
ith
η
per
on-
ing
lue
cal
ion

for

N

959

ica-

3

0

rch,
ever
ose.
ion.
and
ade,